Host genetic factors underlying the link between the microbiome and rheumatoid arthritis

Wells, Philippa

Awarding institution:
King's College London

Host genetic factors underlying the link between the microbiome and rheumatoid arthritis

Philippa M. Wells

A thesis presented for the degree of

Doctor of Philosophy

Department of Twin Research & Genetic Epidemiology

King's College London

United Kingdom

2020

Abstract

Rheumatoid arthritis (RA) is a chronic autoimmune disease, for which presently available treatments are unsatisfactory; a substantial proportion of patients continue to experience symptoms in addition to adverse treatment side-effects. Understanding of the relationship between the commensal microbiota and human health has developed exponentially in recent years, secondary to the establishment of next generation sequencing of the bacterial genome. It is now clear that there is intricate crosstalk between microbiota and human host, affording quite extensive influence on health and disease. Overall, human physiology cannot be adequately understood in the absence of consideration of the microbiome. There is much interest in the link between the microbiome and RA - a primary influence of the microbiota is host immune development and maintenance - and a rich story has been uncovered. However, disentangling of the complex link between the commensal microbiota and RA presents a challenge.

Primarily, epidemiological studies of the RA microbiome are needed to indicate the disease associated microbiota differences, the driving factors behind these, and the temporal trajectory. Only through understanding of these factors can we have a basis from which to link the microbiota with RA aetiology. Prospectively, modulation of the commensal microbiota may be applicable to prevention, and treatment of RA.

In this thesis I apply novel epidemiological approaches to investigate the link between the microbiome and RA, with a focus on host genetic factors. I first consider appropriate methods for pre-processing of 16S sequence data which informs analysis and interpretation throughout the thesis and apply these methods to develop a core dataset for use in further work. To capture genetic risk of RA in Twins UK, I construct a polygenic risk score (PRS), using UK Biobank to additionally confirm efficacy. I calculate taxonomic heritability of the of the gut (stool) and oral (saliva) microbiota, demonstrating substantial heritability of both. This provides an evidence base for subsequent studies in which I investigate the association

between RA PRS and the commensal microbiota. I demonstrate that *Prevotella* within the gut microbiota, which has to date received by far the most attention within the field for it's relevance to RA aetiology, is associated with host genotype. I successfully validate this finding in a separate cohort. The study identified genotype as a factor in key aspects of microbiota composition in RA, the pervasive nature of which lends extrapolation to temporal trajectory; this work indicates that microbiota differences are present before onset of clinical RA. In this instance, the commensal microbiota may potentially be a feasible target for prevention of progression from arthralgia to clinical RA. The study of the oral (saliva) microbiota was less conclusive as to association with genotype. However I made advancements in understanding of factors which influence the oral (saliva) microbiota, understanding of which is much less developed compared to the gut microbiota. In particular, I demonstrate that saliva microbiota are substantially heritable, and diversity is associated with age after controlling for other factors including general health. I move on to apply the converse approach, of investigating microbiota differences whilst controlling for host genotype within a disease discordant twin study design – a powerful method afforded through working with a twin cohort: TwinsUK. Despite practical limitations of recruitment of RA discordant twins, this is the largest study of the gut microbiome of established (treated) RA patients to date. Whilst host genotype and other important factors including diet were controlled for within the study, there were detectable differences in the microbiota. Species level identification of taxa associated with RA was not possible, and interpretation is therefore challenging. Tentatively, these results supported the hypothesis that taxa which are found in higher abundance in control participants in prior studies of the RA microbiome, are associated with treated RA, and therefore a portion of the efficacy of RA medication may be attributed to the microbiome. This paves the way for future studies which are underway to Interrogate of the relationship between the microbiota and RA medication; there is interest in the link between the microbiota and response to RA treatment.

In summary this thesis provides clarity to understanding of the link between the microbiome and RA aetiology, by providing evidence for the key areas required for the field to progress. The work has relevance beyond the RA microbiome field, in general understanding of factors

influencing the gut and oral (saliva) microbiome, and therefore can inform future studies of wider scope.

Acknowledgements

I first of all would like to thank my first supervisor Claire Steves, for giving me this opportunity, and for being a fantastic role model these past few years. Under your guidance I have learnt so much - I have developed professionally and with your unfaltering support in the face of inevitable challenges I have been able to give my best to the PhD. You understood my strengths and secured opportunities for me for which I am very grateful, it has been a delight to be part of your team. I would like to thank my second supervisor Frances Williams for your guidance in so many aspects of my work, but particularly in honing the clarity of my writing - this will be extremely valuable moving forward. Thank you also for always pushing to me achieve my utmost best.

Thank you to all of my colleagues at the DTR, without whom this work would not have been possible. In particular I would like to thank Ruth Bowyer for her incredible patience and kindness (in general!) but particularly in the unenviable task of assisting me with learning R, and Helena Wells for the same but particularly for showing me the ropes when I was new and making me welcome, and both for your friendship. I've been so lucky to meet so many great people here - there's too many to name - but all are excellent colleagues, and many have become friends.

Thank you to our overseas collaborators, particularly Daniel Sprockett at Stanford, and in Switzerland, Axel Finckh, Till Strowig, Robin Lesker and their teams.

Thank you to everyone else in my life who has taken the PhD journey alongside me. In particular my sister Lilly for her quiet understanding and love, and my friends who have really been incredible influences in this time, particularly Simon, Hannah, Danny, and Steve. I'm grateful to my Uncle Rainald for igniting my passion for science with all of his imaginative science-themed gifts to me as a child. I absolutely must also thank my Exeter supervisor who mentored me in my pre-PhD phase, Crawford Winlove. All ofthese people have made this thesis possible.

Publications & Conference Presentations Associated with this Thesis

Original Research Articles

Wells, P.*, Doores, K.*, Couvreur, S., Martin Martinez, R., Seow, J., Graham, C., Acors, S., Kouphou, N., Neil, S., Tedder, R., Matos, P., Poulton, K., Jose Lista, M., Dickenson, R., Sertkaya, H., Maguire, T., Scourfield, E., Bowyer, R., Hart, D., O'Byrne, A., Steele, K., Hemmings, O., Rosadas, C., McClure, M., Capedevila-Pujol, J., wolf, J., Ourseilin, S., Brown, M., Malim, M., Spector, T. and Steves, C., 2020. Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England. *MedRxiv. Under review with Lancet Infectious Diseases.*

Wells, P., Adebayo, A., Bowyer, R., Freidin, M., Finckh, A., Strowig, T., Lesker, T., Alpizar-Rodriguez, D., Gilbert, B., Kirkham, B., Cope, A., Steves, C*. and Williams, F*., 2020. Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *The Lancet Rheumatology*, 2(7), pp.e418-e427.

Semi Zouiouich, Erikka Loftfield, Inge Huybrechts, Vivian Viallon, Panayiotis Louca, Emily Vogtmann, **Philippa M Wells**, Claire J Steves , Karl-Heinz Herzig, Cristina Menni, Marjo-Riitta Jarvelin, Rashmi Sinha, Marc J Gunter, 2020. Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies. *Under review with Microbiome.*

Le Roy, C., **Wells, P.**, Si, J., Raes, J., Bell, J. and Spector, T., 2020. Red Wine Consumption Associated With Increased Gut Microbiota α-Diversity in 3 Independent Cohorts. *Gastroenterology*, 158(1), pp.270-272.e2.

Fabre, M., McKerrell, T., Zwiebel, M., Vijayabaskar, M., Park, N., **Wells, P.**, Rad, R., Deloukas, P., Small, K., Steves, C. and Vassiliou, G., 2020. Concordance for clonal hematopoiesis is limited in elderly twins. *Blood*, 135(4), pp.269-273.

Wells, P., Williams, F., Matey-Hernandez, M., Menni, C. and Steves, C., 2019. 'RA and the microbiome: do host genetic factors provide the link?. *Journal of Autoimmunity*, 99, pp.104-115.

AS Adebayo, G Ackermann, RC Bowyer, **P M Wells**, G Humphreys, R Knight, TD Spector, CJ Steves. (2019). The urinary tract microbiome in older women exhibits host genetics and environmental influences. *BiorXiv.* Accepted/in press with Cell-Host-Microbe.

Nag, A., Kurushima, Y., Bowyer, R., **Wells, P.**, Weiss, S., Pietzner, M., Kocher, T., Raffler, J., Völker, U., Mangino, M., Spector, T., Milburn, M., Kastenmüller, G., Mohney, R., Suhre, K., Menni, C. and Steves, C., 2020. Genome-wide scan identifies novel genetic loci regulating salivary metabolite levels. *Human Molecular Genetics*, 29(5), pp.864-875.

Bowyer, R., Schillereff, D., Jackson, M., Le Roy, C., **Wells, P.**, Spector, T. and Steves, C., 2020. Associations between UK tap water and gut microbiota composition suggest the gut microbiome as a potential mediator of health differences linked to water quality. *Science of The Total Environment*, p.139.

Ni, Y., Yu, G., Chen, H., Deng, Y., **Wells, P.**, Steves, C., Ju, F. and Fu, J., 2020. M2IA: a web server for microbiome and metabolome integrative analysis. *Bioinformatics*, 36(11), pp.3493-3498.

Ruth Bowyer, Nitin Shivappa, **Philippa M Wells**, Tim D Spector, Ailsa A Welch, James Hébert, Claire J Steves, 2020. Update of use of dietary indices to control for diet in human gut microbiota studies. *BiorXiv. Accepted/In Press with Microbiome*.

Conference Presentations

6th-10th October 2019: Keystone, Microbiome: Therapeutic Implications, Kerry, Ireland:- Short Talk & Poster

25th-27th November 2018: Genotype-Phenotype-Fitness, Exeter:- Poster.

12th-14th September 2018: Musculoskeletal and Neurodegenerative disorders Symposium, Trelleborg, Sweden:- Poster.

Contents

List of Figures

List of Tables

List of Abbreviations

ACPA	Anti-citrullinated protein antibody
AD	Alzheimer's disease
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
APC	Antigen presenting cell
ASV	Amplicon sequence variant
Aβ	Amyloid beta
BBB	Blood brain barrier
CCL20	Chemokine (C-C motif) ligand 20
CD4	Cluster of differentiation 4
CRP	C reactive protein
CTLA-4	Cytotoxic T lymphocyte antigen 4
DAMPS	Damage associated molecular patterns
DMARDS	Disease modifying anti-rheumatic drugs
EULAR	European League against Rheumatoid Arthritis
FCR-γ	Fragment crystallisable receptor gamma subunit
FGF	Fibroblast growth factor
FLS	Fibroblast-like synoviocytes
FMT	Faecal Microbiota Transfer
GALT	Gut associated lymphoid tissue
GI	Gastro-intestinal
GPCR	G protein coupled receptor
HLA	Human leukocyte antigen
IBD	Inflammatory bowel diseases
IEC	Intestinal epithelial cell
IFN	interferon
IgG	Immunoglobulin G
IL	Interleukin
iNKT	Invariant natural killer T
JAK	Janus kinase

Kgp	Lyseine gingipain
Kgp	Lyseine gingipain
MAF	Minor allele frequency
MALT	Mucosa associated lymphoid tissue
MHC-II	Major histocompatibility complex class II
MMP	Matrix metalloproteinase
MUC-II	Mucin II
NLR	NOD-Like receptor
NOD1	Nucleotide binding oligomerization domain-containing protein 1
NSAIDS	Non-steroidal anti-inflammatory drugs
OMV	Outer membrane vesicle
OR	Odds ratio
OTU	Operational taxonomic unit
PAD	Peptidyl arginine deiminase
PAMPS	Pathogen associated molecular patterns
PAR2	Protease-activated receptor 2
PCOS	Polycystic ovary syndrome
PDGF	Platelet derived growth factor
PRR	Pattern recognition receptor
PTPN22	Protein tyrosine phosphatase, non-receptor type 22
RA	Rheumatoid arthritis
RANK	Receptor activator of nuclear factor-κB
RANKL	Receptor activator of nuclear factor-κB ligand
RF	Rheumatoid factor
RgpA	Arginine gingipain A
RgpB	Arginine gingipain B
SD	Standard deviation
SE	Shared epitope
SNP	Single nucleotide polymorphism
STAT4	Signal transducer and activator of transcription 4
TGF-β	Transgenic growth factor beta
Th1	T helper cell type 1

Th17	T helper cell type 17
Th2	T helper cell type 2
TLR	Toll-like receptor
TNF-α	Tumour necrosis factor alpha
Tregs	Regulatory T cells
VEGF	Vascular endothelial growth factor

Chapter 1

Introduction part 1

1.1 Thesis Overview

There is mounting interest in the commensal microbiota as a factor in the aetiology of rheumatoid arthritis (RA). The potential influence of host genetic factors has been little explored. In this thesis, I investigate the link between the microbiome and RA, with particular regard to the influence of host genetic factors.

I begin by synthesising and reviewing prior relevant literature. I next consider appropriate methods for analysis of the microbiome. I consider in particular detail, the method for pre-processing the raw sequencing data, which is influential for the resolution of all further analyses in this thesis. I use polygenic risk scoring to capture participants' genetic risk of RA. I take a novel approach of exploring microbiota association with genetic risk of RA in unaffected participants, and juxtapose this by investigating the difference in the microbiota between RA discordant twins, thereby focussing on genetic association in the first approach and controlling for genetic factors in the second.

In this chapter, I introduce the subject areas for the thesis – rheumatoid arthritis and the microbiome. I begin by describing the clinical features, pathophysiology, and treatment of rheumatoid arthritis. I then introduce the microbiome, in the context of human physiology. I consider examples of links between the microbiome and health and disease. In Chapter 2: Introduction 2, I present a literature review in which I consider in detail the current evidence for a link between the microbiome and rheumatoid arthritis, through the lens of host genetics. Thesis aims are presented in Chapter 2.

1.2 Rheumatoid arthritis

1.2.1 Background

Rheumatoid arthritis (RA) is a systemic autoimmune condition affecting 1% of the population, conferring life altering disability, and reduced life expectancy (Guo et al., 2018). It is characterised by synovial inflammation, cartilage and bone destruction, angiogenesis, and production of autoantibodies – anti-citrullinated protein antibody (ACPA) and rheumatoid factor (RF). In addition to articular involvement there are systemic effects including cardiovascular, psychological, renal, pulmonary, ocular, and musculoskeletal pathology (Firestein and McInnes 2017). RA confers life-altering disability for those affected. RA affects both younger and older people, particularly women, and has a typical age of onset between 30 and 80 years (Mueller et al. 2014).

Despite considerable advances in treatment in the last two decades, there remain unmet needs with regards to the treatment of RA. In particular, current therapeutic options show varying efficacy between patients, with some patients being fully or partially unresponsive to disease modifying anti-rheumatic drugs (DMARDs). Complete remission is achieved only in the minority of patients, and instead is maintained via ongoing pharmacological management (Guo et al. 2018). Further, there are a lack of reliable biomarkers for prognosis, therapeutic response, or toxicity.

Herein lies the motivation to further understand rheumatoid arthritis and particularly the implication of modifiable risk factors such as the microbiome; there is a requirement for advancement of understanding regarding the aetiology of RA, to facilitate improvement in preventive measures and clinical management.

1.2.2 Clinical features of rheumatoid arthritis

RA patients present with arthralgia and associated joint swelling. Rheumatoid arthritis typically affects the hands initially, and particularly the metacarpophalangeal and proximal interphalangeal joints in a symmetrical fashion. Other joints – wrist, elbow, ankle, and knee - may also be affected and over time characteristic RA hand deformities may develop (swan neck, boutonniere, Z deformity of the thumb and ulnar deviation).

There is extensive extra-articular involvement in RA, secondary to sequalae of systemic inflammation. Cardiovascular implications are the most common, in particular there is an increased incidence of stroke and myocardial infarction, secondary to aggravation of atherosclerosis, which leads to the shorter life expectancy associated with RA. Renal implications are characterised by an increase in circulating C reactive protein (CRP), Hepcidin and incidence of anaemia. Neurological complications are an increase in incidence of fatigue and depression. There are wider musculoskeletal complications in addition to the classical development of bony erosions which is described below in Section 1.1.4. Osteopenia may develop, with potential to progress to osteoporosis. In muscle cells, insulin resistance occurs leading to muscle weakness.

1.2.3 Aetiology of rheumatoid arthritis

RA is a heterogeneous condition and the detailed aetiology has not been clearly characterised. RA has a strong genetic component and is estimated to be 65% heritable (Kim et al. 2017). Current understanding of the genetic aetiology of RA is discussed in detail in Chapter 2. Genetic factors are necessary, but not sufficient, for disease development. Additional factors are also required: smoking, diet, hormone fluctuation (there is a higher prevalence in women) and more recently suggested the mucosal microbiota. RA disease is subclassified into subsets: seronegative RA, and seropositive RA in

which circulating RA IgG autoantibodies - anti-citrullinated protein antibody (ACPA) and rheumatoid factor (RF) - are present. ACPA is also present in around 1% of the RA unaffected general population, however these individuals are understood to be on the pathway to RA development; ACPA is specific for RA (Finckh et al. 2020).

In RA there is a move to a pro-inflammatory phenotype, in which there is systemic, chronic low-grade inflammation involving both innate and adaptive immune system components, with insidious disease development over decades, finally reaching the fulfilment stage in which classical RA pathological features present. Systemic autoimmunity, evidenced by circulating RA autoantibodies ACPA and less specifically RF, can be detectable years before onset of full clinical RA. There is incremental development of the pathology, prior to the point at which symptoms are evident secondary to inflammation of the synovium, bone and cartilage erosion, and angiogenesis. The phases of RA development are summarised in Figure 1.1.

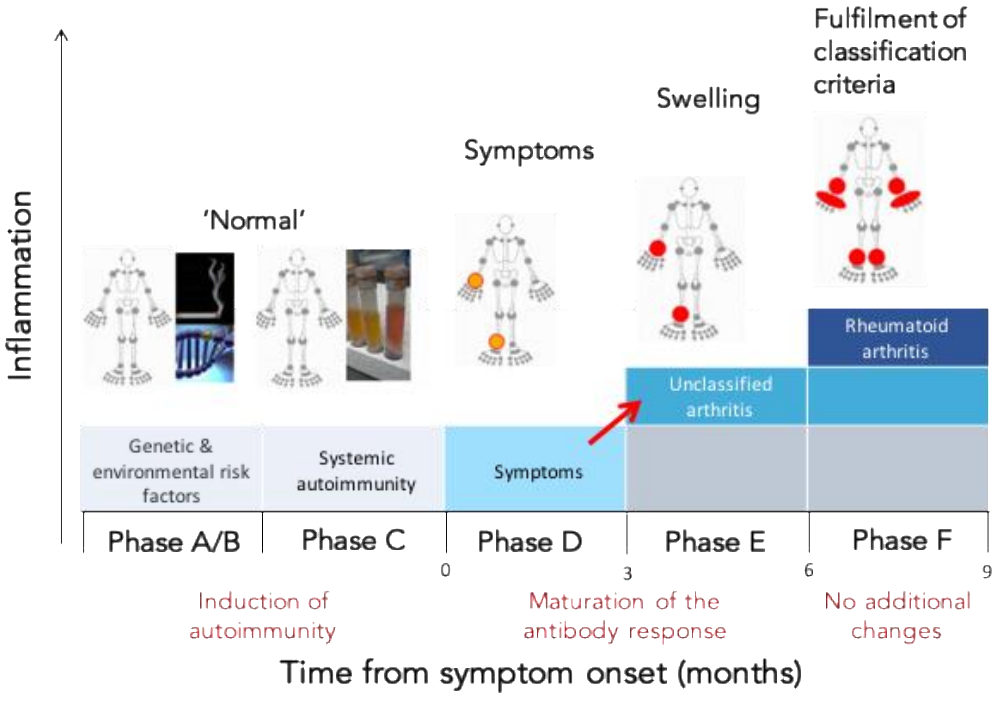


Figure 1.1 The natural history of rheumatoid arthritis
Adapted from Raza K. et al. (2012)

There is interplay between the strongest genetic risk factor - HLA-DR4 shared epitope (SE) and the strongest environmental risk factor - smoking - which leads to development of autoimmunity and production of ACPA auto-antibodies, a cardinal feature of RA. As mentioned prior, smoking leads to post translational modification of peptides within the lung mucosa - specifically citrullination - the conversion of arginine to citrulline. Concurrently, HLA-DR4 SE RA risk alleles encode for amino acids in the binding site of major histocompatibility complex class II (MHC-II), and alter the tertiary structure of the peptide in such a way as to confer higher affinity for citrullinated protein antigens. Other factors leading to protein citrullination in RA are peptidyl arginine deiminase (PAD4) positivity, commensal microbiota (discussed in Section 1.3) and dietary factors (Guo et al. 2018). Citrullinated proteins bound to MHC-II on antigen presenting cells bind CD4+ B and T cells, leading to production of ACPA and T cell activation, respectively. ACPAs promote loss of tolerance to self-antigens, form immune complexes which promote inflammation, and contribute to osteoclastogenesis leading to articular bone degradation and development of bony erosions (Figure 1.2).

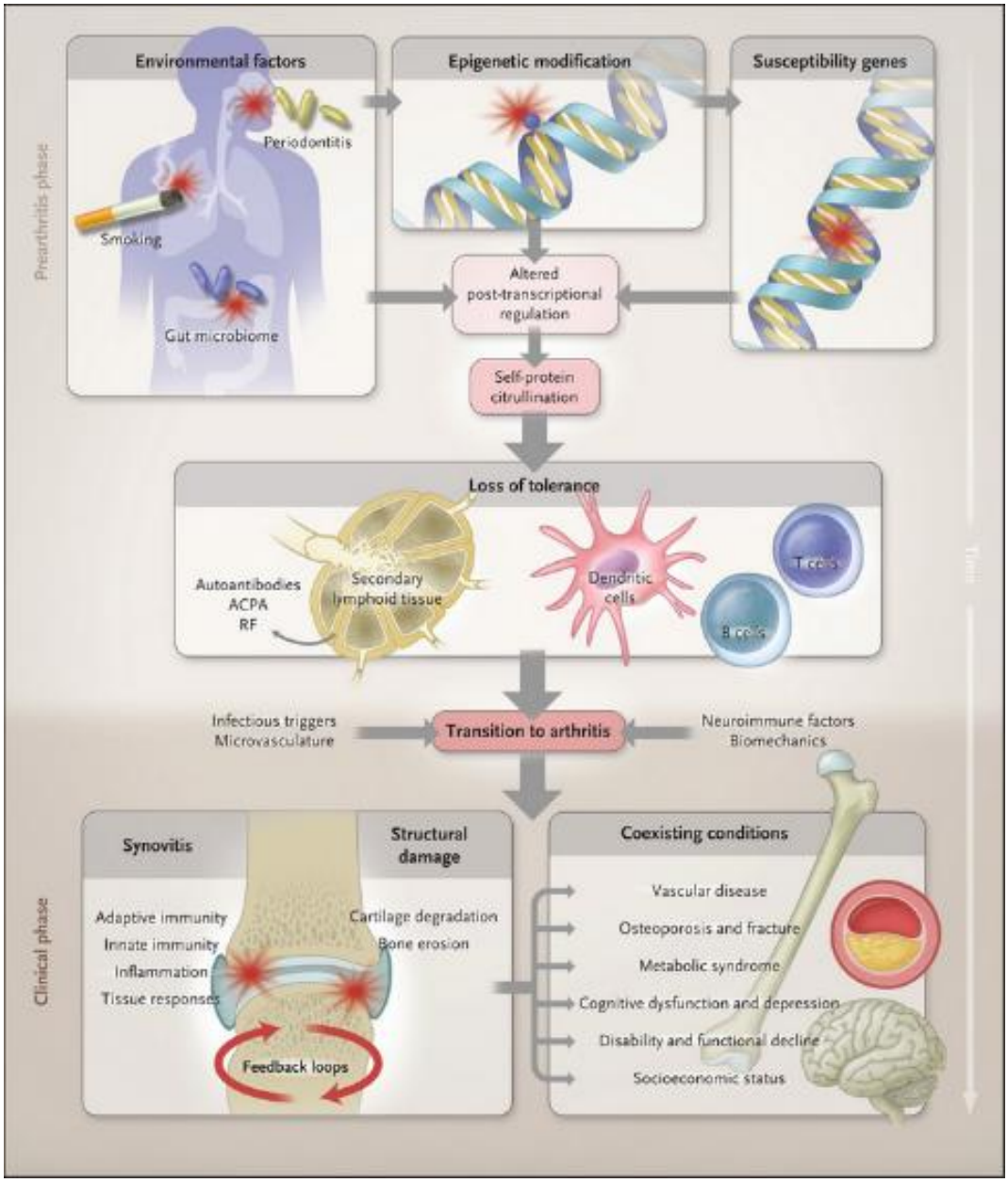


Figure 1.2 Suggested model of RA aetiological development
Adapted from McInnes and Schett (2011)

It is hypothesised that an additional event is required to move aberrant RA characteristic immune activity into the threshold of disease. This is sometimes referred to as the "two hit hypothesis". Triggers suggested include injury causing mechanical insult, infection, neuro-immune factors or dysfunction of the microvasculature (Figure 1.2). These events may trigger the immunological cascade seen in clinical RA, and in particular may trigger articular involvement (McInnes and Schett 2017).

1.2.4 Immunopathogenesis of rheumatoid arthritis

As RA is a heterogeneous condition, the detailed immuno-pathogenesis has not been demonstrated. However common key immune cells involved have been established; a broad framework of the cell types and aetiological cascade beginning from the stage at which synovial inflammation occurs is as follows. Within the synovium, macrophages secrete pro-inflammatory cytokines including tumour necrosis factor alpha (TNF-α), interleukin 1 (IL-1) and interleukin 6 (IL-6). Cytokines stimulate fibroblast-like synoviocytes (FLS), cells comprising the membrane of the synovium, which when activated begin to proliferate, and together with pro-inflammatory cytokines, stimulate receptor activator of nuclear factor-κB ligand (RANKL) expression, and osteoclastogenesis, leading to bone erosion. Activated FLS also secrete proteases, which degrade cartilage. T cells comprise 50% of immune cells in the inflamed joint and pro-inflammatory Th17 cells secrete IL-17, leading to downstream macrophage activation in addition to further stimulation of FLS and expression of RANKL. Activated FLS can migrate to other joints, proliferating pathology at other sites (Figure 1.3).

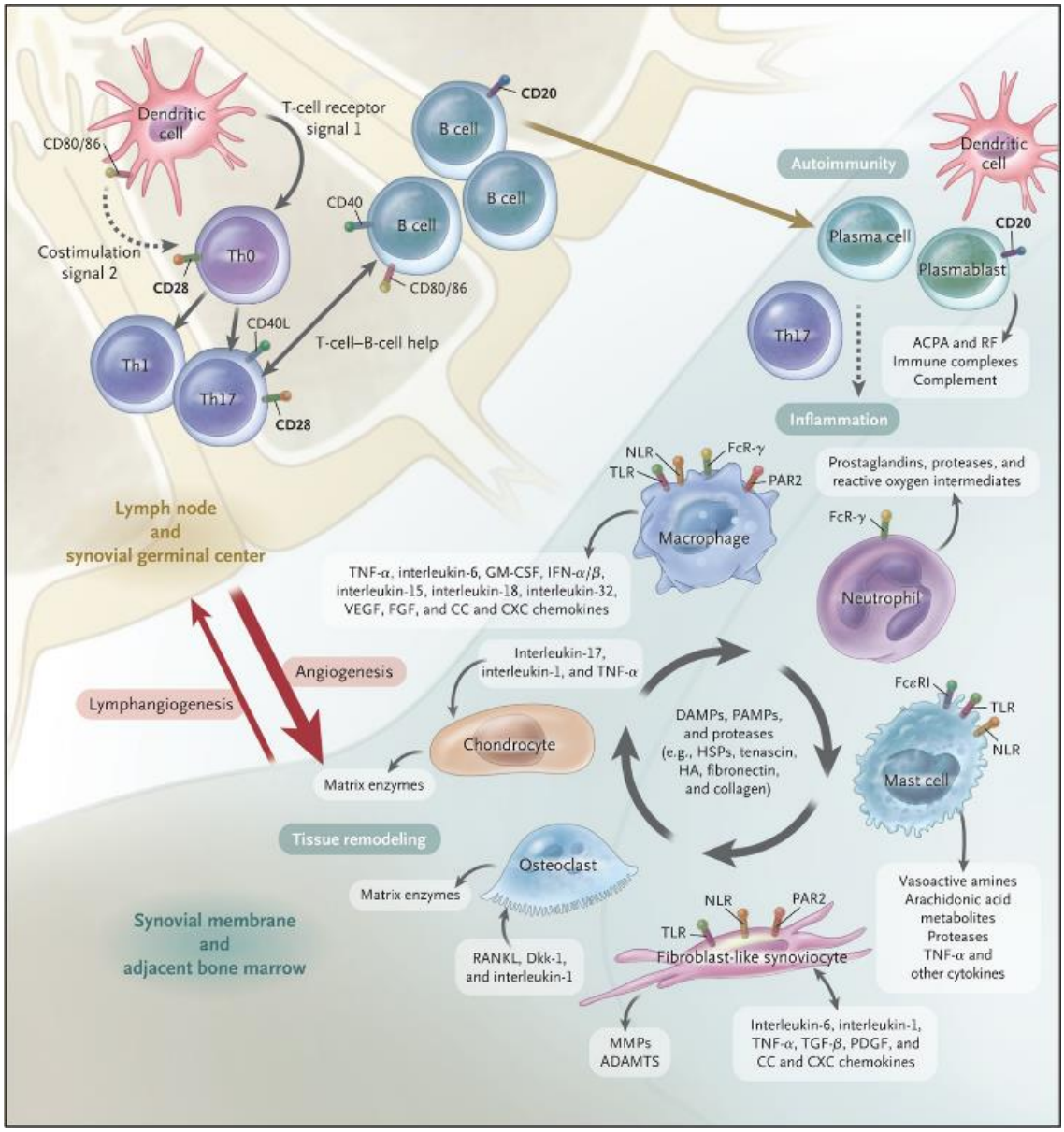


Figure 1.3 Articular immunopathogenesis of RA
Adapted from McInnes and Schett (2011)

Positive feed-back loops drive the pathology forward. Inflammation drives further inflammation, in the context of suppression of regulatory immune elements.

1.2.5 Therapeutic Regime for Rheumatoid Arthritis

Rheumatoid arthritis is primarily managed using a collection of therapeutics termed disease modifying antirheumatic drugs (DMARDs), comprised of synthetic and antibody based

(biologic) drugs. Available DMARDs and their mechanism of action, and key known adverse effects are summarised in Table 1.1. However first-line treatment with the aim of initial pain relief and mediation of inflammation is via non-steroidal anti-inflammatory drugs (NSAIDs) (Bullock et al. 2019). Corticosteroids are more potent anti-inflammatories compared to NSAIDs and may be used systemically (usually by mouth or intra-muscularly) or given intra-articularly in the short term to manage a single painful joint. However, the side effects of corticosteroids are recognised to be too severe to allow their inclusion in a long-term treatment regime. Rapid progression to treatment with DMARDs is usually required, and of these, methotrexate is the first-line treatment.

The European League against Rheumatoid Arthritis (EULAR) recently released updated recommendations for the management protocol for RA (Smolen et al. 2020). After clinical diagnosis of RA, methotrexate is to be started in combination with a glucocorticoid. If methotrexate is contraindicated, leflunomide or sulphasalazine are recommended instead. If improvement is apparent at 3 months and target treatment achieved at 6 months, this regime is continued with dose reduction throughout sustained remission. If the target is not achieved, it must next be established if there are poor prognostic factors for the patient. Poor prognostic factors comprise RF or ACPA seropositivity at high levels; high disease activity score; early joint damage and failure of 2 or more synthetic DMARDS. If poor prognostic factors are present, addition of a biologic DMARD or JAK inhibitor is indicated. Following another 6 months of this regime, if the target is achieved then (as above) the treatment combination can be continued at a dose reduced in intervals to achieve sustained remission. If after 6 months the target is not achieved, different biological DMARDs may be tried, until treatment is successful and remission achieved, at which point the dose can be reduced as above. If poor prognostic factors are absent, a second synthetic DMARD can be added.

In RA research, there is a focus on the pre-clinical stages of the disease – treating the pathology before it is fully established is likely to be the most effective option (Cope 2017).

Indeed, the microbiome may provide a means to abate the pathology in the pre-clinical stages.

DMARD Classification	Name	Mechanism	Known Adverse Effects
Synthetic	Methotrexate	Folic acid inhibitor	Increased liver enzymes, pulmonary damage.
Synthetic	Leflunomide	Pyrimidine synthesis inhibitor	Hypertension, diarrhoea and nausea, hepatotoxicity.
Synthetic	Sulfasalazine	Anti-inflammatory/ immunosuppression	Gastrointestinal, central nervous system, and hematologic adverse effect.
Synthetic	Hydroxychloroquine	Immunomodulation	Gastrointestinal tract, skin, central nervous system adverse effect and retinal toxicity
Biologic	Infliximab	TNF-α inhibitor	Infection (pneumonia and atypical tuberculosis) injection-site reaction.
Biologic	Adalimumab	TNF-α inhibitor	Hypertension
Biologic	Etanercept	TNF-α inhibitor	Severe anaphylactoid transfusion reaction.
Biologic	Golimumab	TNF-α inhibitor	
Biologic	Certolizumab Pegol	TNF-α inhibitor	
Biologic	Rituximab	B cell depletion	Infection, hypertension, hypogammaglobinaemia, viral reactivation, vaccination responses.
Biologic	Abatacept	CD28/CTLA4 system	Infection, malignancy
Biologic	Tocilizumab	IL-6 inhibition	Infections (most notably skin and soft tissue), increases in serum cholesterol, transient decreases in neutrophil count and abnormal liver function.
Biologic	Anakinra	IL-1 inhibition	Injection site reactions, infections, neutropenia, malignancy
Biologic	Tofacitinib	JAK1 &JAK3 inhibition	Zoster infection
Biologic	Baricitinib	JAK1 & JAK2 inhibition	Zoster infection

Table 1.1 Pharmacological therapies used in treatment of RA
Adapted from Guo *et al.* (2018)

1.3 The human microbiome

The human microbiome is the collective term for the community of commensal micro-organisms which co-exist within the human body and encompasses their collective genome and the host habitat. The term microbiota refers more specifically to the collection of micro-organisms which constitute the microbiome (Marchesi et al. 2016). Bacteria constitute a substantial portion of the microbiota and are the most characterised, however viruses, archaea, protozoa and fungi are also present and likely hold important roles (Nishida et al. 2018, Sadanand 2019).

Traditionally, bacteria have been viewed as an opponent to health, and instigators of disease. However, it is now understood that the relationship is infinitely more complex. In conceptualising human medicine, commensal microbiota are integral: they are integral to diverse aspects of development and physiological function, and thus health and disease. The microbiota are required for the development of a normal immune system and continue to shape the immune state throughout life (Kabat et al. 2014). They produce up to one third of metabolites detectable in blood including signalling molecules (Menni et al. 2017). They are capable of post-translational peptide modification (Konig et al. 2015), induction of morphological change and alteration of mucosal cell connectivity (Lin et al. 2015). There is therefore scope for the microbiome to influence diverse pathologies – not only those in which inflammation plays a central role. In turn, factors relating to the human host influence the composition of the commensal microbiota: age, diet, xenobiotics, geographical environment, genetic factors, neonatal factors, and disease state. The bi-directional relationship between microbiota and host is summarized in Figure 1.4.

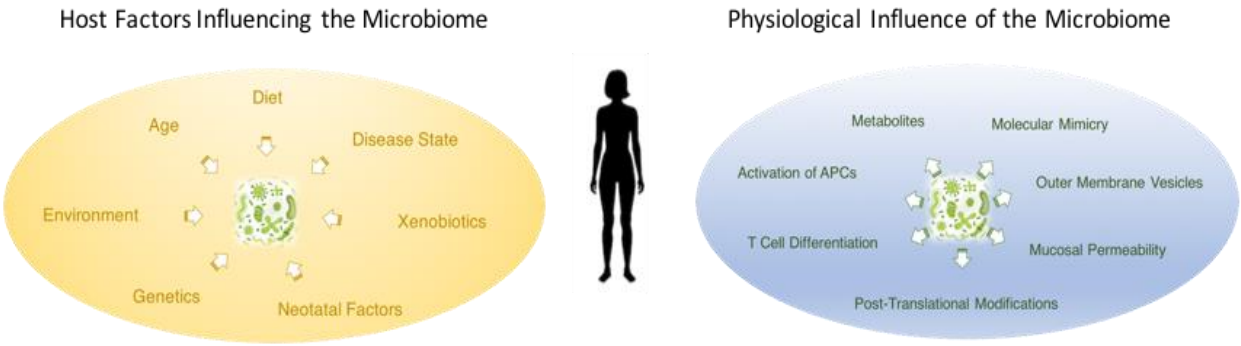


Figure 1.4 Interaction between human host and commensal microbiota

Whilst there is some limited evidence of a pre-natal microbiome, the present paradigm considers that seeding of the microbiome occurs at birth. In babies delivered vaginally, this occurs via ingestion of maternal faecal material and other bio-fluids (Combellick et al. 2018). Influence of host genotype on the microbiome occurs via a variety of factors: the immune and metabolomic microbial environment, dietary preferences and disease state, and determination of biological sex. However, diet is the most influential known determinate of microbiota composition, providing substrates which lead to propagation of the oral and gut microbes which comprise the majority of the commensal microbiota (Singh et al. 2017). Ingestion of xenobiotics such as drugs and synthetic vitamins alters the composition of the microbiota, and accounts for some of the microbiota association with disease (Carmody and Turnbaugh 2014, Vila et al. 2020). Difference in microbiota composition compared to the general population is also a generalised feature of disease. In particular, a lower alpha (within sample) diversity of the gut microbiota is a frequent feature of a diseased state (Jackson et al. 2018a).

The microbiota in turn influence host physiology in a manner that is dependent on the above influencing factors. In particular, the microbiota metabolise nutrients and xenobiotics (Vila et al. 2020) and microbial metabolites are a key mechanism via which the microbiota

influence host physiology (Visconti et al. 2019). The microbiota are an important determinate of the host immune state via numerous mechanisms such as activation of antigen presenting cells (APCs), Th17 cell differentiation and influencing mucosal permeability via intestinal epithelial cell (IEC) tight junctions (discussed in Sections 1.3.1, 1.3.2 and 1.3.3). Commensal microbiota are also capable of post translational modification of host peptides, leading to differential immune response (Konig et al. 2015). A more novel suggested mechanism by which the microbiota may influence host physiology is via outer membrane vesicles (OMVs), which are the bacterial equivalents of lysosomes. OMVs are ubiquitously produced by gram negative bacteria, they function as a transport mechanism for bacterial peptides within the host and maintain the microbial ecosystem. It was recently demonstrated that intestinal epithelial cells are able to uptake OMVs via endocytosis, and also that OMVs can transmigrate through IECs (Jones et al. 2020).

1.3.1 The commensal microbiota and the immune system

An important mode of influence of the microbiota on host physiology is via the immune system. The microbiome is integral to the induction, training and homeostasis of the host immune system (Lazar et al. 2018). This relationship is particularly prominent with the gut microbiota, which constitutes the majority of the commensal microbial load, and has the closest interaction with the immune system via the gut associated lymphoid tissue (GALT), which is a sub-category of the mucosa associated lymphoid tissue (MALT). However, microbiota at other sites have important immune relevance (Colombo et al. 2016). Throughout life the commensal microbiota inform both local tissue specific, and systemic immune tone, achieved through complex interactions between intestinal epithelial cells (IECs), microbiota derivatives and leukocytes (Kabat et al. 2014). Therefore, the host microbiota relationship may provide a mutual pin-hold underlying chronic inflammatory conditions.

1.3.2 The commensal microbiota and innate immunity

The largest interface between host and microbiota comprises the epithelial cells of the intestinal tract which in healthy conditions form a physical barrier, enforced by numerous intercellular tight junctions. The intestinal epithelial cells are a frontline site of microbial monitoring, secondary to compartmentalised expression of toll like receptors (TLRs). TLRs belong to the pattern recognition receptors (PRRs) family (Németh et al. 2016a, p. 9). Ligands for these receptors are pathogen associated molecular pattern molecules (PAMPs), damage-associated molecular pattern molecules (DAMPs) and microbe associated molecular pattern molecules (MAMPs), with the latter being characteristic of commensal microbiota. TLRs detect derivatives of potentially harmful pathogens, leading to innate and adaptive immune activation. Conversely, TLRs function to inhibit the inflammatory cascade which may otherwise be triggered by presence of commensal microbiota. TLR expression by IEC shows a polarized pattern, with expression at the basolateral surface, thereby minimising the inflammatory response during healthy conditions. TLR2 and TLR4 are predominantly expressed within the small intestine, whilst TLR5 is expressed on the basolateral surface of colonic IECs (Hug et al. 2018). Commensal microbiota also contribute to the maintenance of the integrity of the IEC barrier in infection or adverse conditions, by inducing anti-inflammatory cytokines and anti-apoptotic signalling cascades (Kabat et al. 2014).

Specialised goblet cells of the IEC barrier synthesise mucus and produce a thick mucus layer which covers the apical surface of the intestinal epithelium. This 150 μm thick layer serves as a protective barrier and is augmented with IEC derived anti-microbial peptides (AMPs). Structurally, the mucus layer is organised around a hyper-glycosylated mucin MUC2 (Shan et al. 2013). The colonic mucus layer comprises two sub-layers: a dense inner layer that is relatively sterile, and a looser outer later in which commensal microbiota reside (Kabat et al. 2014).

The mucus layer is not simply a structural barrier, but it is also molecularly involved in the induction of immune tolerance to commensals, particularly via the innate immune system (Figure 1.5). In addition to bactericidal properties, AMPs modulate cell function via chemotaxis, TLR signalling and during wound healing (Kabat et al. 2014). The mucus layer encourages a more balanced commensal colonisation, and MUC2 is capable of mitigation of dendritic cell mediated inflammatory response (Shan et al. 2013). Whilst some AMPs such as α-defensin are constitutively expressed by IECs, others are regulated by IEC sensing of pathogen-associated molecular patterns (PAMPs). The commensal microbiota inform IEC mediated immune tissue development of isolated lymphoid follicles. Peptidoglycan derived from the cell wall of commensal bacteria activates the cytoplasmic PRR NOD1 in IEC and induces b-defensin 3 and CCL20 secretion, which drive formation of isolated lymphoid follicles.

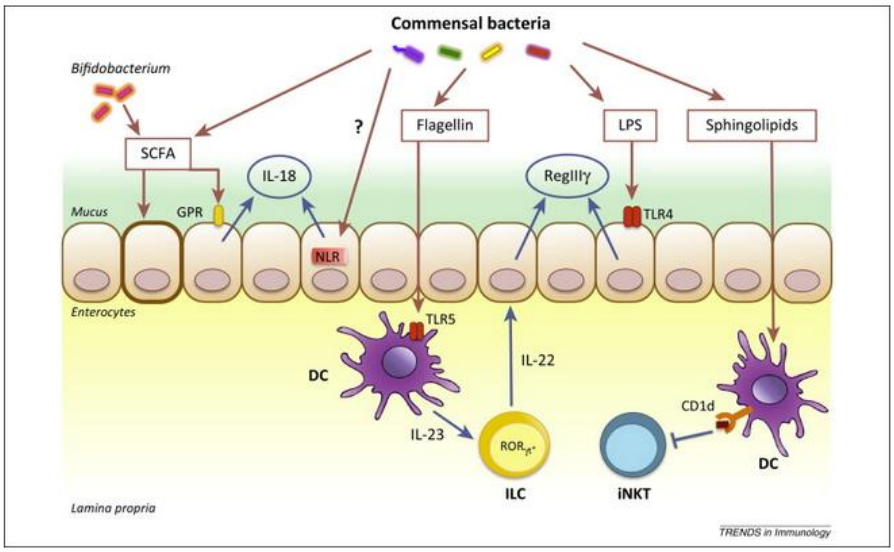


Figure 1.5 Examples of commensal microbiota influence on innate immune responses
Adapted from Kabat *et al.* (2014)

1.3.3 The commensal microbiota and adaptive immunity

The commensal microbiota are also widely influential in the recruitment, activation and differentiation of adaptive immune cells (Figure 1.6). Overall, the commensal microbiota influences the local immune milieu to promote immune tolerance by limiting pro-inflammatory Th1 and Th17 responses, promotion of Th2 responses and differentiation and function of regulatory T cells (Tregs). Examples of interactions between the commensal microbiota and adaptive immunity are as follows. Signals from commensal microbiota induce production of B cell activating factor of the TNF family (BAFF), a proliferation–inducing ligand (APRIL) and transforming growth factor β (TGFβ) in IEC and dendritic cells (DC), which in turn promotes the differentiation of B cells into immunoglobulin (Ig) A^+ plasma cells (Goto 2019). Following activation by commensal bacteria, follicular dendritic cells (FDC), which are the primary source of TGF-β in Payer's patches, also promote the differentiation of B cells into IgA^+ plasma cell. The intestinal microbiota can regulate function of the innate lymphoid cells (ILC), which in turn promote T cell-independent IgA induction through the interaction of membrane bound lymphotoxin with DC. The soluble form of ILC-derived lymphotoxin supports T cell-dependent IgA induction by promoting homing of T cells to the lamina propria, thereby influencing the T follicular helper (Tfh) population (Gutzeit et al. 2014). Polysaccharide A (PSA) produced by *Bacteroides fragilis* directly promotes Treg cell differentiation directly via TLR2 and indirectly by conditioning DCs. Microbiota-derived short-chain fatty acids (SCFA), a by-product of metabolism, may promote Treg cell generation either directly through signalling via G protein-coupled receptor (GPCR) 43 or indirectly via IEC. *Clostridium* species induce TGF-β production in IEC, which promotes Treg cell differentiation in the colon (Bauché and Marie 2017).

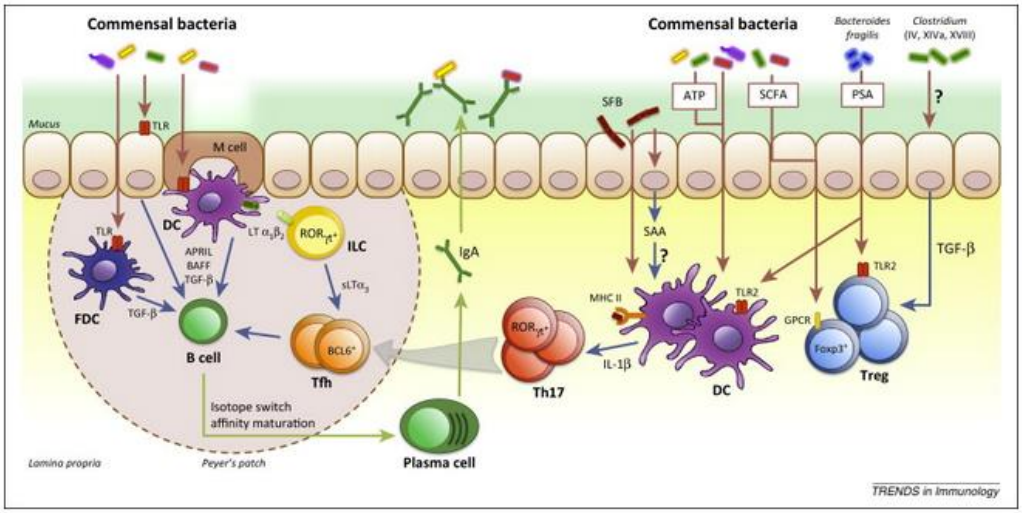


Figure 1.6 Examples of commensal microbiota influence on the adaptive immune system
Adapted from Kabat *et al.* (2014)

Overall, when operating optimally, this immune system–microbiota alliance allows the induction of protective responses to pathogens and the maintenance of regulatory pathways involved in the maintenance of tolerance to innocuous antigens (Belkaid and Harrison 2017). However, there is a hypothesis that facets of modern living in Westernised countries, particularly the overuse of antibiotics, banishing of helminths and move to highly processed and lower fibre diet have selected for a microbiota that lack the characteristics required to establish balanced immune responses (Rodriguez-Castaño et al. 2017). This phenomenon is proposed to account for some of the dramatic rise in autoimmune and inflammatory disorders in parts of the world (Broussard and Devkota 2016).

1.3.4 The oral microbiome: environmental niches

Microbiota composition is secondary to its host environment, for example the oral cavity, skin, or intestinal tract, and there is therefore substantial variation between body sites. Tissue-specific microbiota have differing associations with disease. The human microbiome project (HMP), which represents a ground-breaking effort to characterise the healthy human microbiome, has investigated the microbial population of 5 major body areas – oral

cavity, intestinal tract, nasal cavity, vagina and skin. Of these, the oral microbiome is one of the most diverse; the oral cavity provides at least 9 distinct environmental niches – subgingival sulcus, supragingival tooth surface, gingiva, hard palate, buccal mucosa, tongue, saliva, throat and tonsils (Figure 1.7).

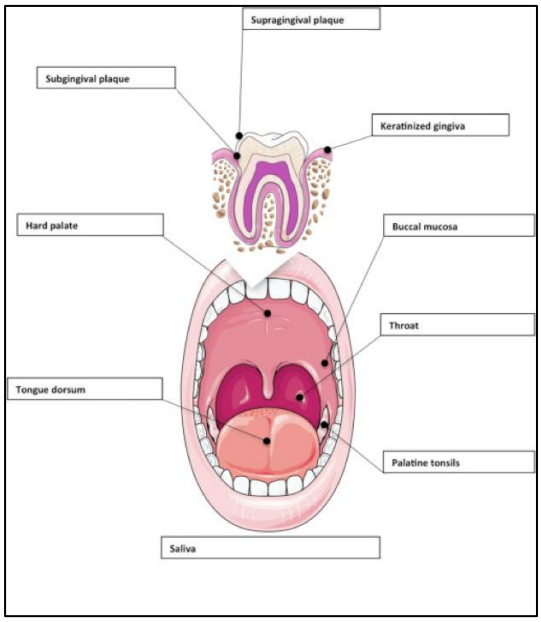


Figure 1.7 Environmental niches of the oral cavity
Adapted from Sampaio-Maia *et al.* (2016)

Within the oral microbiota there are 150 genera identified comprised of 700 species of which 51% are officially named, 13% are unnamed but cultivated and 34% of which are known only as uncultivated phylotypes (Chen et al. 2010). Most habitats within the oral cavity are dominated by *Streptococcus,* followed in abundance by *Haemophilus* in the buccal mucosa, *Actinomyces* in the supragingival plaque and *Prevotella* within subgingival plaque.

In contrast to the lower gastro-intestinal tract, the oral cavity is well suited to the formation of biofilms, having far lower motility and a lack of a mucus layer. Congregation of bacteria within biofilms allows a greater array of interaction between bacteria, such as horizontal gene transfer, and interspecies functional modulation (Alseth et al. 2019).

1.3.5 The oral microbiome: physiological influence

The oral microbiota have an important physiological role in nitric oxide homeostasis. Nitric oxide has diverse functions and is a facilitator of vasodilation, nerve impulse transmission, host defence mechanisms and cellular energetics. The classical pathway for nitric oxide generation is via nitric oxide synthases. However, an alternative pathway also having an important role is the nitrate to nitrite to nitric oxide pathway. Here, nitrate present in food such as leafy green vegetables is sequestered by the salivary glands and released in saliva, where conversion to nitrite is facilitated by the oral microbiota which use these molecules in respiration as the final electron acceptors in the electron transport chain. Key oral bacteria identified as having a role in nitrate reduction are *Streptococcus salivarus*, *Streptococcus mitis, Streptococcus bovis, Veionella spp., Staphylococcus epidermis, Staphylococcus aureus, Nocordia spp* and *Corynebacterium* (Hezel and Weitzberg 2015).

1.3.6 The oral microbiome examples of pathophysiological influence outside the mouth: Alzheimer's disease

Understanding the microbiome and the nuances of its influence on disease is a task of considerable proportions, and one which we are only beginning to address.

To date there are numerous disease associations with the oral microbiome, and the oral microbiota is increasingly appreciated as having important pathyophysiological systemic consequences (He et al. 2015, Gao et al. 2018). Within the oral cavity, diseases in which the oral microbiota are causally implicated are caries, periodontal disease of the tooth supporting periodontium, oral mucosal diseases, oral cancer and peri-implantitis.

The systemic diseases associated with altered oral microbiota include obesity, rheumatoid arthritis (Chapter 2; Chapter 7), HIV infection, liver cirrhosis, inflammatory bowel diseases (IBD), polycystic ovary syndrome (PCOS), diabetes, atherosclerosis and cardiovascular disease and more recently Alzheimer's disease. A frequent feature underlying many conditions is chronic low-grade inflammation, however there are also numerous more disease-specific mechanisms.

In genetically susceptible individuals, periodontal disease is initiated by the oral microbiota and particularly the dental biofilm formation which constitutes plaque. Chronic periodontal disease is characterised by gingival swelling preceding erosion of the gingiva, bone, and ligaments, leading to tooth loss as the supporting structures disintegrate. Chronic periodontitis is classified as general chronic periodontitis when more than 10 teeth are lost, and localised periodontitis below this threshold. In both instances there is systemic inflammatory involvement (Kinane et al. 2017). Risk factors for periodontal disease include genetic predisposition to systemic inflammation (having overlap with RA genetic risk factors, discussed in Chapter 2), smoking, diabetes mellitis, low educational attainment and psycho-social variables (Kinane et al., 2017; Figure 1.8).

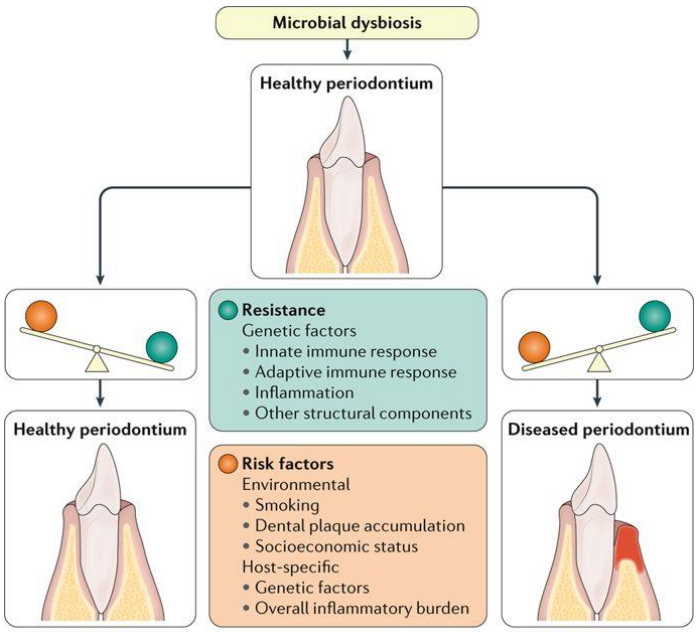


Figure 1.8 Susceptibility factors for periodontal diseases
Adapted from Kinane *et al.* (2017).

Periodontal disease has been historically associated with a group of bacteria known as the 'red complex', which comprises *Porphyromonas gingivalis*, *Treponema denticola* and *Bacteroides forsythus,* which were reported as particularly associated with clinical measures of periodontal disease – dental pocket depth and gingival bleeding on probing (Socransky et al. 1998). It has since been advanced that wider microbiota associations or specific gene combinations thereof converge to shape and stabilise a disease provoking microbiota. The detailed aetiology and microbiota involvement is yet to be characterised, but disease development is likely to be secondary to presence of a pathogenic oral microbiota in the context of other pre-disposing risk factors and genetic propensity for inappropriate immune response.

Recently the commensal microbiota more generally have been implicated in Alzheimer's disease (Dominy et al. 2019) and also in the cognitive decline that precedes it. Longitudinal studies have shown that periodontal disease is associated with declining cognitive function and may be a contributory factor (Pritchard et al. 2017). A recent retrospective cohort study showed that a 10-year exposure to periodontal disease was associated with subsequent AD (Chen et al. 2017). This is in accordance with understanding that chronic low-grade inflammation is a risk factor for AD. Whilst PD is associated chronologically with AD, there is some question as to the direction of effect - a two-way relationship is likely.

Alzheimer's disease (AD) is the most common form of dementia, accounting for 60-70% of cases. AD is characterised by an insidious and progressive decline in cognitive ability which confers loss of independent functioning (Fernández et al. 2010). The pathological hallmarks of AD are extracellular amyloid beta (Aβ) plaques and intracellular tau neurofibrillary tangles (NFTs) within the brain (Emery et al. 2017). More recently microglial activation has been identified as a hallmark of Alzheimer's disease and inflammatory mechanisms associated with microglial activation may play a pivotal role in neurotoxicity. Tau is a microtubule associated protein, which in AD is hyperphosphorylated within neurons and accumulates in the neuronal cytosol. The mechanism of tau

hyperphosphorylation in AD is yet to be determined, however, evidence suggests that neuronal cell stress responses to pro-inflammatory cytokines may be important. Aβ is a 4 kDa peptide, formed by the sequential cleavage of amyloid precursor protein (APP), a large transmembrane protein, by beta secretase1 (BACE1) and gamma secretase (Emery et al. 2017). There are various isoforms of Aβ, varying by length (Aβ40, Aβ42), solubility and post translational modification. Insoluble Aβ accumulates to form AD characteristic extracellular plaques. The level of soluble Aβ42 oligomers has been demonstrated to correlate most closely with clinical severity (Condic et al. 2014). The neurotoxicity of Aβ in AD is secondary to multiple mechanisms including oxidative stress, Aβ associated receptor mediated mitochondrial dysfunction and excitotoxicity (Picone et al. 2014, Cheignon et al. 2018).

Porphyromonas *gingivalis* (*P.gingivalis),* a key periodontal disease bacterium, which is also associated with RA (Chapter 2; Chapter 7) has been associated with Alzheimer's disease. *P.gingivalis* has been demonstrated to be present in the saliva and cerebrospinal fluid (CSF) of probable AD patients with mild to moderate cognitive impairment. Virulence factors produced specifically by *P.gingivalis* known as gingipains – cysteine proteases which are transported by OMVs, have been found in the brains of AD patients. Specifically, within the hippocampus of deceased RA patients, arginine gingipain B (RgpB) co-localised with neurons and astrocytes (but not microglia), and was found to be intracellular. RgpB co-localised with phosphorylated and un-phosphorylated tau, and with intracellular soluble Aβ42. *In vitro* cultured neuronal cells expressing high molecular weight tau demonstrated tau truncation in response to the addition of both *P.gingivalis* and gingipains - RgpB and lyseine gingipain (Kgp). Tau truncation and fragmentation is a likely precursor to generation of insoluble and hyperphosphorylated tau. Tau truncation was abated by addition of a cysteine protease inhibitor but not by broad spectrum antibiotics, indicating that the reaction is mediated by gingipain virulence factors of *P.gingivalis* as opposed to other bacterial peptides. In murine models, after oral infection with *P.gingivalis* every other day for six weeks, the *P.gingivalis* 16s rRNA gene was detectable in the brain suggesting bacterial translocation. Presence of *P.gingivalis* in the brain was associated with levels of

Aβ42 (Dominy et al. 2019). Further supporting this link, Aβ is an anti-microbial peptide and can be considered a component of the innate immune system (Pritchard et al. 2017). This places the potential link with translocated oral microbiota and microbial peptides feasibly early in the disease pathogenesis. In this scenario, oral microbiota are likely to contribute to microglial activation – microglia are inflammatory mediator immune cells of the brain, activation of which are now identified as a third hallmark of AD in addition to Aβ and tau pathology. A suggested sequence of events is oral microbe translocation, Aβ deposition and glial cell activation, but substantial further work is needed to shed light on the association. There is a link here with RA; quite extenstive work on the oral microbiota and RA has focussed on *P.gingivalis* and associated gingipains (discussed in Chapter 7).

Overall, the oral microbiota are of great interest having systemic influence and also relevance in an array of clinical conditions. With regards to rheumatoid arthritis, the oral microbiota may provide the link between RA and one of the primary RA risk factors periodontal disease. This is discussed in detail in the literature review in Chapter 2, and further in Chapter 7: Genetic risk for RA and association with the oral microbiota. Regarding research studies, the oral microbiota holds a methodological advantage in that it is readily sampled: microbiota from each of the oral sites are easily obtained. In contrast, microbiota sampling of the gastro-intestinal tract represents a substantial methodological challenge (see Section 1.4.4).

1.4 The gut microbiome composition and example of physiological and pathophysiological influence

The term 'gut microbiome' refers to the microbiome of the gastro-intestinal (GI) tract. The GI tract encompasses the oral cavity, oesophagus, stomach, duodenum, small intestine, cecum, large intestine, colon and rectum (Figure 1.9). However, 'gut microbiome' is widely used to refer to the colonic microbiome.

1.4.1 The gut microbiome: environmental niches

Within the colon, it is estimated that hundreds of thousands of bacterial species co-habit (Almeida et al. 2019). The GI tract is the largest interface between the internal and external environment. The gut (colon) microbiota constitute the majority of the commensal microbiota. Over 99% of the gut microbiota belong to the four phyla Actinobacteria, Proteobacteria, Firmicutes and Bacteroidetes, of which the latter two predominate (Nishida et al. 2018, Almeida et al. 2019). As is observed with the oral microbiome, the gastro-intestinal tract represents an array of microbial niches and environmental variation, which confers variation in microbiota composition along the GI tract (**Figure 1.9**).

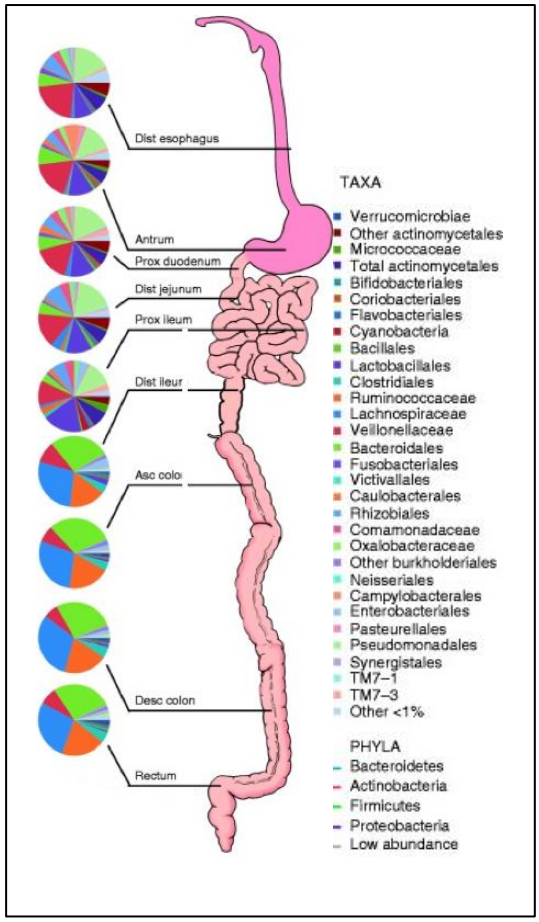


Figure 1.9 Overview of microbiota variation along the GI tract
Adapted from Vuik et al. (2019)

The most marked shift in microbiota composition along the GI tract occurs between the distal ilium and the colon and is secondary to a gradient shift in pH, oxygen concentration and the abundance of anti-microbial peptides among other factors (Hall et al. 2017). In addition to compositional changes, there is marked variation in micro-organism abundance along the GI tract. Microbiota abundance in bacterial cells per gram going down the gut within the lumen is estimated at an order of 10^{1} in the stomach, increasing to 10^{3} in the duodenum, 10^{4} in the jejunum, 10^{7} in the ileum and finally 10^{12} in the colon reflecting a more hospitable microbial environment (Shah et al. 2020).

1.4.2 The gut microbiome: physiological influence

Gut bacteria are key regulators of digestion along the gastrointestinal tract; commensal bacteria play an important role in the extraction, synthesis, and absorption of many nutrients and metabolites, including short-chain fatty acids (SCFA), vitamins, bile acids, lipids and amino acids. These factors are in addition to those described above in Section 1.2.1 in regulating the development, homeostasis, and function of innate and adaptive immune cells.

SCFA, the most abundant of which are acetate, propionate and butyrate hold diverse physiological roles. Butyrate is integral to GI function as it is the primary metabolic substrate for colonic epithelial cells, and it also holds wider roles via epigenetic function having histone modification capability thus altering gene expression and induction of apoptosis of colonic cancer cells (Nishida et al. 2018, Rowland et al. 2018). Butyrate and propionate have been demonstrated to influence gluconeogenesis conferring a beneficial role in glucose homeostasis. Propionate is suggested to be important for achieving satiety. It is a substrate for intestinal G protein – coupled receptors (GPCRs) which are also known as fatty acid receptors. Propionate is produced typically by *Bacteroides* species and

particularly *Negativicutes*, and also *Clostridium* species. Acetate supports the microbial community as it is an essential co-factor for the growth of other bacteria. For example, *Faecalibacterium prausnitzi* (*F.prau*), a bacterium uniformly associated with health, requires acetate. *F.prau* is a key producer of butyrate alongside other members of the Firmicutes phylum. Acetate also has a physiological role in cholesterol metabolism and lipogenesis. Acetate is the most abundant SCFA and is a metabolite produced by many taxa. The gut microbiota are the primary source of circulating serotonin neurotransmitter production. In total the gut microbiota produce up to one third of the metabolites detectable in the bloodstream, which enter the systemic circulation largely via the hepatic portal vein (Sharon et al. 2014).

The commensal gut microbiota have an important role in vitamin synthesis. It is long established that gut microbiota are responsible for synthesis of vitamin K and act as the primary source. Vitamin K is required for the function of the clotting cascade. In addition, the gut microbiota are capable of synthesis of B group vitamins – biotin, cobalamin, folates, nicotinic acid, pantothenic acid, pyridoxine, riboflavin, and thiamine. A recent study showed that 40-65% of gut microbiota taxa are capable of synthesis of all eight B vitamins (Magnúsdóttir et al. 2015). It was previously identified that bacteria belonging to the Bacteroidetes phylum hold around 90% of B vitamin synthesis capability. B vitamins of microbial source are primarily utilised across the gut microbiota, contributing to microbiota community stability, and therefore contributing indirectly to host homeostasis.

Another function of the gut microbiota is proteolysis of proteins from dietary sources and endogenous proteins from enzymes, mucins and sloughed intestinal cells. This produces shorter peptides, singular amino acids, fatty acids, and gases. Conversely, gut microbiota synthesis of polypeptides from free amino acids contributes to amino acid bioavailability (Ma and Ma 2019).

Bile acid composition and abundance is heavily influenced by the gut microbiota. In addition to their classical role in lipid digestion and absorption in the small intestine, and cholesterol

homeostasis, bile acids also function as hormones which are implicated in a variety of metabolic processes (Staels and Fonseca 2009). Bile acids have systemic effects – through reabsorption they are transported via the circulatory system to act as ligands for bile acid receptors which are expressed in multiple cell types throughout the body. Bile acids are considered an important class of bioactive molecule. Within the intestine, the commensal microbiota are determinants of bile acid metabolism secondary to bacterial excretion of bile salt hydrolases. Synthesis of the primary bile acids cholic acid and chenodeoxycholic acid occurs in the liver, and these are converted to deoxycholic acid and lithocholic acid, respectively, by bile salt hydrolases produced by the gut microbiota. Gram positive bacteria and, in particular *Lactobacillus*, *Bifidobacterium*, *Clostridium* and *Enterococcus,* have been identified as major producers of bile salt hydrolases. In turn, bile acids influence bacterial proliferation, and therefore can be considered mediators of host-microbiome crosstalk (Foley et al. 2019).

Commensal gut microbiota have a crucial role in impeding pathogenic bacteria colonization by inhibiting their growth, for example by consuming available nutrients and producing bactericidins, and successful competition against pathogens via nutrient metabolism, pH modification, antimicrobial peptide secretion and effects on cell signalling pathways. Gut microbiota also prevent bacteria invasion by maintaining the integrity of intestinal epithelium tight junctions. These factors are in addition to those described above regarding regulating the development, homeostasis, and function of innate and adaptive immune cells.

1.4.3 The gut microbiome: example of pathophysiological influence - Inflammatory Bowel Disease

IBD, encompassing Crohn's and ulcerative colitis (UC), are chronic relapsing and remitting conditions in which there are varying degrees of intestinal inflammatory damage with extra-intestinal complications. Symptoms include stomach pain, bloating, fatigue and fever.

Patients may show signs of blood in the stool, altered bowel habit, weight loss and anaemia. A pathophysiological link with the microbiota is not surprising: GI immune tone, colonocyte mucus barrier and tight junction defects are centrally implicated in the pathology (Ungaro et al. 2017). The pathogenesis of IBD is not clearly characterised, however a leading hypothesis is that in genetically susceptible individuals there is an abnormal immune response to commensal gut microbiota, following exposure to multiple environmental factors. UC, in particular, has been associated with differences in the gut microbiota compared to controls, and particularly an expansion of facultative anaerobes of the Enterobacteraceae family (Zhu et al. 2018, Lo Presti et al. 2019). There is aetiological crossover between IBD and RA, including shared genetic susceptibility factors and as such it is not uncommon for IBD patients to develop an IBD arthropathy which is RA-reminiscent. The joint pathology of IBD patients who also have arthropathy tends to differ from that of RA, with affectation primarily of the large joints – knees, ankles, wrists, elbows and hips (Orchard 2012). Faecal microbiota transfer (FMT) – the transfer of faeces from healthy donors to IBD patients, has been posited as a promising option for treating IBD (Tian et al. 2019). FMT is already employed for the treatment of *Clostridium difficile* infections and may have much wider use in the future. Nevertheless, this 'blanket' approach carries risks; given our current state of knowledge of the microbiota, there is potential for unforeseen and damaging effects (König et al. 2017). A more targeted approach has been suggested for IBD, specifically, 'precision editing' of the gut microbiota using tunstate. Here, Enterobactereceae which are particularly associated with IBD are metabolically inhibited by tunstate, which replaces the final electron acceptor in the electron transfer chain of these facultative anaerobes. In this way, the electron transport chain in Enterobacteraceae is rendered obsolete, but the wider microbiota are unaffected. Implementation of this method in murine models showed promise, however it has yet to be trialled in humans (Zhu et al. 2018).

1.4.4 The gut microbiome: challenges in sampling

In contrast to the oral microbiota, the gut microbiota represent a substantial challenge with regards to sampling – samples of GI tract tissue are not readily accessible. For this reason, the gut microbiota are routinely measured using the faecal microbiota as a proxy. It is worth noting that the faecal microbiome and the colonic microbiome taken with colonic biopsies are different. For example, it is evident from studies comparing biopsies with faecal samples that taxa with adhesive pili which provide anchorage to the mucosa and epithelia are under-represented in stool compared to the colon (Mora Ortiz 2019). However, as faecal samples are uniformly used for the study of the gut microbiota, because of ease of collection compared to the invasive procedure needed to collect colonic biopsies, the differences between the colon and faecal microbiota should be consistent across studies.

1.5 Conclusion

In conclusion, RA is a heterogeneous condition, in which there is aberrant activity of the immune system. There is a requirement to further understand the condition in order to improve clinical management. The oral and gut commensal microbiota are intimately implicated in many facets of host physiology including, but not limited to immune function. This thesis seeks to examine the relationship between RA and the oral and gut microbiota. It is very timely because recent studies have provided tantalising evidence of this aetiological link. In the following chapter, I critically examine the prior evidence for the link between RA and the microbiome, and present the thesis aims.

Chapter 2

Introduction 2: Literature Review & Thesis Aims

In this chapter I present a review of published evidence with regards to the link between the microbiome and RA, through the lens of host genetic factors. In this work I evaluate the genetic aetiology of RA; the association of the microbiome with RA; the potential influence of host genetic factors in the link between the microbiome and RA. Following this, I present the aims which will be addressed within this thesis.

The literature review is presented as a paper, originally published in the Journal of Autoimmunity in February 2019.

This publication is available at DOI: 10.1016/j.jaut.2019.02.004

2.1 Literature Review

Journal of Autoimmunity 99 (2019) 104–115

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

'RA and the microbiome: do host genetic factors provide the link?

Philippa M. Wells[a,*], Frances M.K. Williams[a], M.L. Matey-Hernandez[a], Cristina Menni[a], Claire J. Steves[a,b]

[a] The Department of Twin Research and Genetic Epidemiology, King's College London, St Thomas' Hospital, Lambeth Palace Road, London, SE1 7EH, UK
[b] Clinical Age Research Unit, Kings College Hospital Foundation Trust, London, UK

ARTICLE INFO

Keywords:
Genetics
Microbiome
Rheumatoid arthritis
Autoimmunity
Prevotella
Amplicon sequence variants

ABSTRACT

Rheumatoid arthritis (RA) is a chronic autoimmune disease, characterised by painful synovium inflammation, bony erosions, immune activation and the circulation of autoantibodies. Despite recent advances in therapeutics enabling disease suppression, there is a considerable demand for alternative therapeutic strategies as well as optimising those available at present. The relatively low concordance rate between monozygotic twins, 20–30% contrasts with heritability estimates of ~65%, indicating a substantive role of other risk factors in RA pathogenesis. There is established evidence that RA has an infective component to its aetiology. More recently, differences in the commensal microbiota in RA compared to controls have been identified. Studies have shown that the gut, oral and lung microbiota is different in new onset treatment naïve, and established RA patients, compared to controls. Key taxonomic associations are an increase in abundance of *Porphyromonas gingivalis* and *Prevotella copri* in RA patients, compared to healthy controls. Host genetics may provide the link between disease and the microbiome. Genetic influence may be mediated by the host immune system; a differential response to RA associated taxa is suggested. The gut microbiome contains elements which are as much as 30% heritable. A better understanding of the influence of host genetics will shed light onto the role of the microbiome in RA. Here we review the role of the microbiome in RA through the lens of host genetics, and consider future research areas addressing microbiome study design and bioinformatics approaches.

1. Introduction

Rheumatoid arthritis (RA) is a chronic multisystem autoimmune condition, characterised by painful swelling of the synovial joints bone and tendon damage. RA affects almost 1% of the population and is highly debilitating, with a profound effect on life quality in both young and elderly people. It has a significant economic impact on society partly through loss of working ability [1]. Despite considerable recent advances in therapeutics enabling disease suppression, there is still a sizeable minority of patients where drug therapies are ineffective or poorly tolerated, with potential serious adverse effects and the need for regular blood monitoring to early detect these [2]. Thus, there is a considerable demand for alternative therapeutic strategies as well as optimising those available at present.

RA is a common complex disease derived from the interplay between genetic and environmental factors [3,4]. Known risk factors include periodontal disease [5], smoking [4], diet [6] and hormone fluctuation – the disease is more prevalent in women [5–7]. The relatively low concordance rate between monozygotic twins, at 20–30%, contrasts with high heritability estimates of ~65% [8], indicating a substantive role of other risk factors in RA pathogenesis in genetically susceptible individuals [9]. There is early evidence of epigenetic influence [10] and a longstanding appreciation of the possible role of infection as triggering the immune activity [11]. Despite genome-wide association studies (GWAS) and GWAS meta-analyses of increasingly large samples, the 349 variants identified for RA account for less than a third of the estimated heritability – a common finding in complex traits. This "missing heritability", i.e. the inability to account for the proportion of phenotypic variance contributed by genetic factors – may be due to (i) structural variation (such as copy number variations-CNVs), (ii) by rare variants, or (iii) by environmental factors which are influenced by host genetics.

RA presents disease subsets, the clearest of which are characterised by the presence or absence of auto-antibodies – seropositive and seronegative RA respectively. There is a genetic aetiological difference, with seropositive RA being secondary to the human leukocyte antigen (HLA) DRB1 (encoding the major histocompatibility complex -MHC) – the shared epitope. Heritability differs by disease subtype, and is

* Corresponding author.
E-mail address: Philippa.Wells@kcl.ac.uk (P.M. Wells).

https://doi.org/10.1016/j.jaut.2019.02.004
Received 12 November 2018; Received in revised form 20 February 2019; Accepted 20 February 2019
Available online 05 March 2019

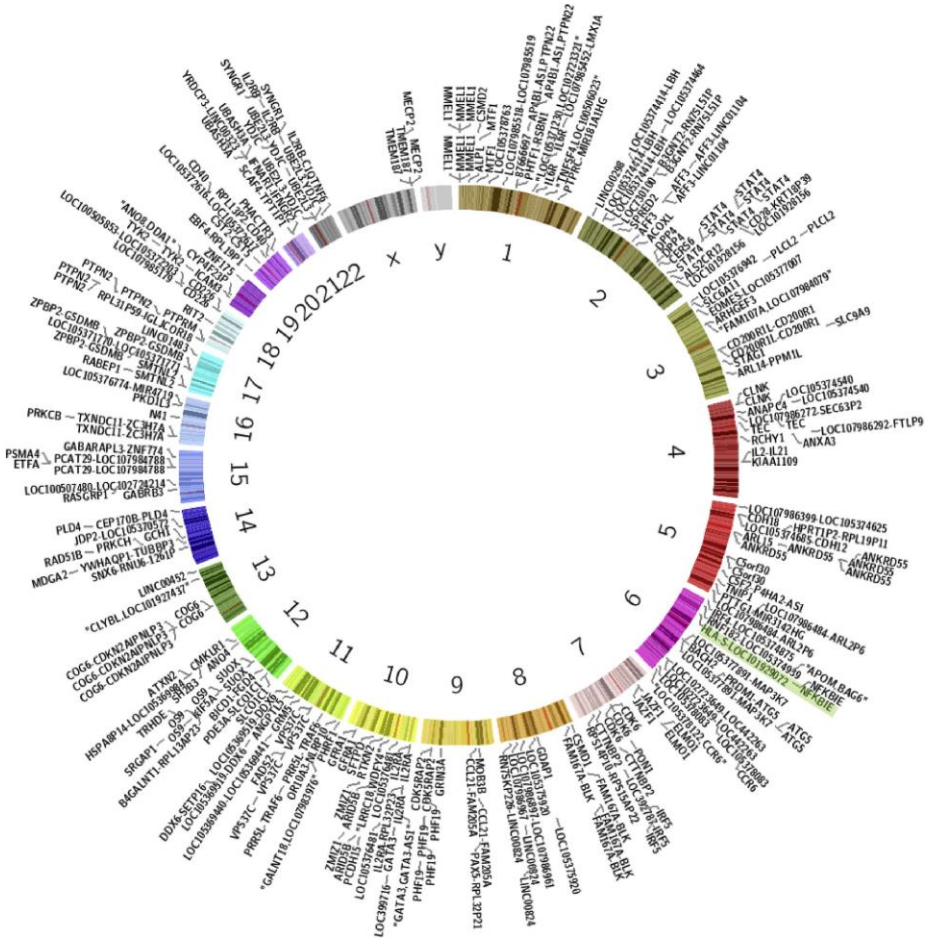


Fig. 1. Genetic Associations with rheumatoid arthritis shown to date. HLA type (highlighted) shows direct mechanistic links with the oral and gut microbiome.

estimated at around 65 and 30% for seropositive and seronegative disease, respectively. Seropositive disease is much more clearly defined so will be the focus of this review. Symptoms develop in the later stages of immune dysfunction following a prolonged period of autoimmunity which, in seropositive RA, is marked by a surge in circulation of autoantibodies – rheumatoid factor and anti-citrullinated protein [6]. Even in seronegative RA the onset of symptoms is typically preceded by years, or even decades, of increased levels of circulating pro-inflammatory cytokines [1]. This has prompted the theory that an environmental factor triggers disease progression. The microbiome, and particularly the alimentary tract microbiome, offers one such candidate mechanism [12].

The microbiome is the collective genome of the vast community of commensal micro-organisms - predominantly bacteria, which inhabit epithelial surfaces [5], blood [13] and tissue [14]. These organisms form a vast symbiotic community, the importance of which in physiological function has only recently been uncovered, secondary to availability of next generation sequencing [15]. The microbiome – the bulk of which is contained within the intestine [16], constitutes a substantial antigenic load and is essential for normal immune development in the neonate, with a life-long role in immune education [17]. It is therefore highly plausible that it has a role – possibly very early in life – in the promotion of subsequent autoimmunity. Further, the microbiome represents a unique physiological interface - a signalling hub, integrating environmental inputs with genetic and immune signals [15].

There is substantial diversity in the microbiome amongst the population [18], and individual differences may be influenced by diet [19], host genetics [20], age [21], host general health [22], antibiotic exposure and bowel habit. Specific variations associate with diseases, and to date the microbiome has been implicated in a vast array of medical conditions, including inflammatory bowel disease [23], cancer

[24] and obesity [25]. The gut microbiome is the most influential, and are a number of ways that it can be altered therapeutically: pro or prebiotics, diet and faecal microbiota transplantation, raising the possibility of hitherto unexplored treatment options.

Studies suggest that the microbiome and particularly the gut microbiome is different in people with RA and may be implicated in the pathogenesis of RA [18,26]. However, the many cross-sectional observations are challenging to place in the context of the temporal evolution of autoimmune disease which evolves over decades and is usually treated as soon as symptoms manifest and the diagnosis is confirmed. In RA there is a gradual transition to a pro-inflammatory phenotype, facilitating the development of disease in genetically susceptible individuals. There may be a bi-directional relationship between RA pathogenesis and the microbiome, in which the microbiome may contribute to the pro-inflammatory phenotype during the propagation stage of autoimmunity. To date nine epidemiological studies [5,18,26,35,44,117–120] have shown alterations in the microbiome in both treatment naïve - excluding an influence of drug effects on the microbiome therefore, and established RA patients, with mechanisms demonstrated in mouse and cellular models. Taxa associations replicate across studies to a greater degree than is usually shown in disease specific microbiome studies, as shown by a recent meta-analysis [27].

Whilst host genetic factors clearly predispose to RA they may also mediate in part the interaction between the microbiome and RA pathogenesis. The gut microbiome itself contains elements which are as much as 30% heritable [20]. Given that the identified genetic risk loci in RA are associated with immune function, it is feasible that risk genotypes for RA act in part via the microbiome. Therefore, the microbiome may explain part of the aforementioned missing heritability in RA; microbes produce a range of enzymes, chemicals, hormones and vitamins that may interact with host metabolism, contributing to as much as one third of the metabolites identifiable in human blood [28].

In this review, the evidence for the role of the gut microbiome in RA will be evaluated, through the lens of host genetic factors. An overview of the genetic aetiology of RA will be given, followed by an evaluation of the evidence for the role of the microbiome in RA, and potential links between host genetics and the microbiome in RA. Areas for further research will then be considered.

1.1. Genetic aetiology of rheumatoid arthritis

In the past few years large-scale human GWAS of RA having over n = 100,000 participants have led to the identification of a substantial number of associated genetic loci. Several common pathways associated with prevalence, severity and progression of rheumatologic disease have been identified. Forty GWAS found a total of 349 SNPs across all chromosomes (Fig. 1), that are associated with RA (Table 1), and 100 SNPs have been replicated on meta-analysis. These studies have revealed that the dominant risk loci in RA are, unsurprisingly for autoimmune disease, located in the MHC region [8] on 6p21.3 and these account for the major proportion of the current heritability explained in RA. Classically, this MHC association is explained mechanistically by the 'shared epitope' hypothesis [8]. The HLA-DRB1 shared epitope alleles - *04:01, *04:04, *04:05, *04:08, *10:01, *01:01 and *01:02 encode for amino acids in position 70 to 74 within the binding site of MHC-II; they therefore influence the host response to extracellular immune ligands. The shared epitope alleles are associated with increased susceptibility to RA, and particularly anti-citrullinated protein antibody (ACPA) positive RA; carriers of these alleles have enhanced immune response to citrullinated proteins. Heterozygosity for human leukocyte alleles (HLA) may confer greater risk of disease due to a wider auto-antigen repertoire of carriers. MHC variants have wider impact than the classic shared epitope hypothesis as this region is characterised by extremely dense and diverse sequence variations, and contains around 250 genes encoding numerous immune molecules in addition to HLA such as complement factors, cytokines and other proteins involved in antigen processing. The complexity of genomic control mechanisms across this region is enormous and only just being unravelled [8,29] (see Fig. 1).

Over 100 associated non-MHC loci have been identified through GWAS meta-analyses [29]. These include variants at the gene encoding protein tyrosine phosphatase, non-receptor type 22 (PTPN22) which functions as a negative regulator of T cell receptors [30], peptidylarginine deiminase 4 (PADI4) which encodes enzymes active in protein citrullination [31], signal transducer and activator of transcription 4 (STAT4) which encodes a transcription factor specific for T cell maturation [32], and TNFAIP3 which encodes tumour necrosis factor alpha (TNFα) [33]. These protein coding genes together influence the immune response and collectively promote a shift to a pro-inflammatory phenotype and increased sensitivity to immune stimulation [34]. Walsh and colleagues characterised the role of protein coding RA associated SNPs and showed involvement in both the innate and adaptive immune systems, which would support a shift to a pro-inflammatory phenotype: jak-STAT signalling pathway; IL-12 mediated signalling; endocytosis; T cell signal transduction; signalling downstream of interleukins and T cell receptors; cytokine signalling; cell adhesion molecules and B lymphocyte cell surface molecules. In keeping with most common complex trait GWAS results, the associated SNPs largely do not reside in protein coding regions, and may act distally with unidentified genes. Non-coding variants are likely to have a role in regulation of immune mediating gene expression [34].

However, as with other complex traits, a full understanding of genetic risk in RA has proved evasive; known risk loci explain only 15% of the estimated heritability, indicating that numerous associations are yet to be discovered [8]. This is perhaps because GWA studies were designed to detect common genetic loci associated with disease, yet genetic risk in RA may be mostly driven by rare variants, with a minor allele frequency (MAF) of less than 0.05 giving accumulated aetiological effect. Research methods must accommodate the polygenic nature of RA, with variable genetic architecture between individuals who may possess numerous RA associated variants of modest effect. Polygenic risk scores (PRS) provide a weighted genetic risk score for individuals, combining individual genotype data with the strength of the disease association for each risk variant. Association of PRS with phenotypes can be modelled, offering an effective option for RA research, and provides a surrogate model of RA allowing examination of the host genetic factors without the confounding influence of the disease or its treatment.

1.2. Gut microbiome in RA

Studies have shown that the gut [18], oral [5], and to a lesser extent, lung [35] microbiomes have been implicated in RA when comparing RA patients to healthy controls (see Table 2). Whether this association is causal has not yet been established. The gut microbiome has been the focus of the RA microbiome link - it constitutes over 80% of the total microbial biomass, with the closest links to the immune system. Nine studies have reported changes in diversity and taxa present in the microbiome of RA patients compared to age, gender and weight matched controls. Whilst lower gut microbiome diversity is known to be a generalised feature of disease [22], the taxonomic and bacterial gene associations with RA are of greater etiological interest. However, there are discrepancies in these associations across the published studies. A link with host genotype, mediated by HLA type in addition to a more general pro-inflammatory genetic predisposition in RA, is suggested.

Prevotella *copri* (P.*copri)* is the most frequently reported bacterial species showing variation of abundance between RA patients and unaffected controls. However, P.*copri* is associated with other inflammatory conditions including metabolic syndrome, insulin resistance, type II diabetes and atherosclerosis [36], in addition to RA [37,38], and may thrive relative to other bacteria within an

Table 1
GWA studies of rheumatoid arthritis.

Study	Cohort			Genetic Association	Mapped Trait
	Ancestry	Cases	Controls		
WTCCC 2007 [81]	European	1860	2938	HLA-DRB1, PTPN22, MHC	RA
Plenge et al., 2007 [82]	European	1522	1850	TRAF1-C5, PTPN22, HLA-DRB1, IL6ST, ANKRD55, SPRED2, C5orf30 PXK, RBPJ, CCR6, IRF5, AFF3, CCL21, IL2RA, CD247, IL2, IL21, SH2B3, BATF, IKZF3, UBASH3A, TNFRSF14, PTPN22, REL, AFF3, STAT4, CTLA4, TNFAIP3, TRAF1, C5, PRKCQ, CD40, POU3F1, KIF3, HLA-DRB1	RA
Plenge et al., 2007 [83]	European	397	1211	TNFAIP3, OLIG3	RA
Liu et al., 2008 [84]	European	89	NR	MAFB, QKI, IFNK, LASS6, CST5, LMO4, CENTD1, PON1	response to TNF antagonist
Julia et al., 2008 [85]	European	400	410	SALL3	RA
Raychaudhuri ét al. 2008 [86]	European	3393	12,460	MMEL1, TNFRSF14, CDK6, CCL21, KIF5A, PIP4K2C, CD40, PRKCQ, PTPN22, HLA-DRB1, TNFIP3, OLIG3	RA
Cui et al., 2009 [87]	European	531	849		ACPA measurement
Gregerson et al., 2009 [88]	European	2418	4504	REL, CTLA4, BLK, PTPN22, TRAF1, C5	RA
Kochi et al., 2010 [89]	Japanese	2303	3380	CCR6, STAT4, TNFAIP3, OLIG3, HLA-DRB1	RA
Stahl et al., 2010 [90]	European	5539	20,169	IL6ST, ANKRD55, C5orf30, PXK, RBPJ, CCR6, IRF5, AFF3, CCL21, IL2RA, CD247, IL2, IL21, SH2B3, BATF, IKZF3, UBASH3A, TNFRSF14, PTPN22, REL,AFF3, STAT4, CTLA4, TNFAIP3, TRAF1, C5, PRKCQ, CD40, POU3F1, KIF3, HLA-DRB1	RA
Padyukov et al., 2010 [91]	European	1147	1853	HLA	RA
Freudenberg et al., 2011 [92]	Korean	801	757	BLK, TRHDE, ARHGEF3, HLA-DRB1, PADI4	RA
Terao et al., 2011 [93]	Japanese	1247	1486	AIRE, PFKL, HLA, PADI4	RA
Hu et al., 2011 [94]	Korean	100	600	APOM	RA
Okada et al., 2012 [95]	Japanese	4074	16,891	B3GNT2, ANXA3, CSF2, CD83, NFKBIE, ARID5B, PDE2A, ARAP1, PLD4, PTPN2, ETS1, FLI1, GCH1, PRKCH, ZNF774, PRKCB1, IRF8	RA
Wang et al., 2012 [96]	European	1157	NR	SPSB1, SLC6A11, ENOX1, MDGA2, ENSG00000102921, PSMA4, RCHY1, EFTA	RA response to drug
Krintel et al., 2012 [97]	European	196	NR	NR2F2, MAP2K6, ALPL, CBLN2, QPCT, CNTNAP4 PDE3A, SLCO1C1	RA, response to TNF antagonist
Myouzen et al., 2012 [98]	Japanese	2303	3380	NFKBIE, RTKN2	RA
Eyre et al., 2012 [99]	European	3297	15,870	HLA, ANKRD55, MMEL1, REL, SPRED2, AFF3, STAT4, CD28, CTLA4, RBPJ, GIN1, TNFAIP3, IRF5, CCL21, TRAF1, IL2RA, DDX6, CD40, PADI4, POU3F1, GATA3, ARID5B, CD5, VPS37C, RASGRP1, TLE3, IRF8, IKZF3, RCAN1, RUNX1, MMEL1, SPRED2, AFF3, TAGAP, IRF5, IRF8, IKZF3, GSDMB, ORMDL3, RCAN1, RUNX1, AFF3, CTLA4, ICOS, RBPJ, IKZF3, GSDMB, ORMDL3, IRAK1, TMEM187, HCFC1, STAT4, SLC9A9, CD28, CTLA4, PTPN22, RSBN1, IL2RB, TYK2, RAVER1, ICAM3, REL, DNASE1L3, PXK, GIN1, C5ORF30, BLK, CCR6, PTPN2, ANKRD55, BACH2, ELMO1, IL6R, AFF3, TRAF1, CTLA4, ICOS, IKZF3, GSDMB, ORMDL3, IRAK1, TMEM187, HCFC1, IL2RB, TYK2, ICAM3, RAVER1, DNASE1L3, PXK, GATA3, GIN1, C5orf30, DDX6, SETP16, TNFSF18.TNFSF4, TAGAP, SPRED2, TNIP1, ANKRD55, CD2, COG6, IL6R, ACTN1	RA
Cui et al., 2013 [100]	European	2706	NR	4 SNPs, genes not reported.	RA response to TNF antagonist
Negi et al., 2013 [101]	North Indian	706	761	ARL15, HLA-DQA2, HLA-DQB1, C6orf10, HLA-DQA1	RA
Okada et al., 2013 [29]	European East Asian	14,361 4873	42,923 17,642	ACOXL, AFF3, ANKRD55, ARID5B, ATG5, ATM, BLK, C1QBP, C4orf52, C5orf30, CL19, CCL21, CCR6, CD2, CD226, CD28, CD40, CD5, CD83, CDK2, CDK4, CDK6, CEP57, CASP8, CFLAR, CLNK, COG6, CTLA4, CXCR5, ABHD6, PXK, DNASE1L3, EOMES, ETS1, FADS1, FADS2, FADS3, FCGR2A, FCRL3, GATA3, GRHL2,HLA-DRB1, IFNGR2, CSF3, IKZF3, IL2, IL21, IL20RB, IL2RA, IL2RB, CSF2, IL3, IL6R, IRAK1, IRF4, IRF5, IRF8, JAZF1, LBH, LOC100506023, LOC145837, LOC339442, MED1, MTF1, INPP5B, NFKBIE, P2RY10, PADI4, PLCL2, AHNAK2, PLD4, PPIL4, PRKCH, PRKCH, PRKCQ, PTPN2, PTPN22, PVT1, RAD51B, RASGRP1, RCAN1, REL, RTKN2, RUNX1, LOC100506403,SFTPD, SH2B3, PTPN11, SPRED2, STAT4, SYNGR1, TAGAP, TEC, TNFAIP3, MMEL1, TNFRSF14, TNFRSF9, TPD52,TRAF1, C5, TRAF6, RAG1, RAG2, TXNDC11, TYK2, YDJC, UBE2L3, WDFY4, ZNF438, PADI4, B3GNT2, CLNK, CSF2, IL3, HLA-DRB1, NFKBIE, AHNAK2, PLD4, IRF8, UBASH3A, P2RY10, LOC339442, FCGR2A, CASP8, CFLAR, ANKRD55, PVT1	RA
Orozco et al., 2014 [102]	European	3034	5271	TNFRSF14	RA
Bossini-Castillo et al., 2014 [103]	European	1148	6008	GRM5, RNASEH2B, FAM124A, CLYBL, MICA, HLA-B, HLA-DRB1, HLA-DQA1, SMIM21	RA
Kim et al., 2014 [104]	Korean European	2234 10,288	7065 35,502	UBASH3A, ETS1, FLI1, TNFSF4, SYNGR1, LBH, COG6, RAD51B, EOMES, MMEL1, PADI4, POU3F1, PTPN22, CD2, IL6R, FCRL3, FCGR2A, PTPRC, REL, SPRED2, AFF3, STAT4, CD28, CTLA4, RPP14, C4orf52, RBPJ, ANKRD55, C5orf30, OLIG3, TNFAIP3, TAGAP, CCR6, IRF5, BLK, CCL21, TRAF1, IR2RA, PRKCQ, GATA3, ARID5B, CD5, DDX6, OS9, RASGRP1, KIF23, TLE3, IRF8, IKZF3, GSDMB, PTPN2, TYK2, CD40, RCAN1, RUNX1, UBE2L3, IL2RB, TNFAIP3	RA

(*continued on next page*)

Table 1 (*continued*)

Study	Cohort			Genetic Association	Mapped Trait
	Ancestry	Cases	Controls		
Senapati et al., 2014 [105]	Indian	281	157	PPM1L, ARL14, PTPRM	response to methotrexate, RA
Jiang et al., 2014 [106]	Han Chinese	952	943	SPRED2, AFF3, CCR6, TRAF1, DPP4, CDK5RAP2, DPP4, MHC HLA-DQA2, HLA-DQB1	RA
Govind et al., 2014 [107]	South African	263	374	LOC100131866, NR5A2, KIAA1542, HLA-DRB1	RA
de Rooy et al., 2015 [108]	Western European	262	NR	HUNK, SCAF4, AP000255.6–001	RA, ACPA, joint damage measurement
Honne et al., 2016 [109]	Japanese	282 moderate 131 severe	31 mild RA	MAP3K7, BACH2, WDR27, GFRA1, CSMD2, SMARCAL1 NLRP10, PKD1L3, IST1, ATXN1L, SMARCAL1, CDH18	RA response to TNF antagonist
Julia et al., 2016 [110]	Spanish	896	282	IRX1	RA, RF
Marquez et al., 2016 [111]	European	3911	10,398	COG6, PTTG1, PTPN22, ATG5, TNFAIP3, TNPO3, IRF5, BLK, IGAM3, TYK2, UBE2L3	RA, SLE
Saxena et al., 2017 [112]	Arab	283	221	HLA-DRB1, CDH6, SMTNL2, GGT6, CD200R1	RA
Bluett et al., 2017 [113]	European	62	175	GJA5, ACP6	response to methotrexate, RA, joint damage measurement
Joo et al., 2017 [114]	Korean	385	326	SLA, TG, SRGAP1	RA
Wei et al., 2017 [115]	European	3323	15,785	ANKRD55, HLA-DQB1	RA disease progression
Yoo et al., 2017 [116]	Korean	120	118	TGFA	

inflammatory host environment [39]. Therefore, individuals with RA risk genotypes, leading to pro-inflammatory immune phenotype, may potentially constitute an ecological niche. This could be in addition to a possible role in disease causation, but needs to be borne in mind when assessing association studies.

Scher and collaborators performed 16s analysis on faecal samples from 144 participants – new onset RA (n = 44), chronic RA (n = 26), psoriatic arthritis (n = 16) and healthy controls (n = 28). They found that in the gut microbiome *P.copri* was most abundant in patients with new-onset RA compared with those with chronic treated RA ($p < 0.01$), psoriatic arthritis ($p < 0.01$) or healthy controls ($p < 0.01$) [26]. This important finding was partially replicated in the study below by Zhang and co-workers [18]. Higher abundance of *P. copri* in the gut microbiome is a characteristic on new-onset RA, they suggest, in which inflammation is relatively unabated by medication. Increased abundance of *P.copri* in new onset RA patients correlated with a decrease of *Bacteroides fragilis* (*B. fragilis*), an important regulator of regulatory T cell (Treg) function. Tregs function in the establishment and maintenance of immune tolerance. This therefore suggested that *P.copri* may influence RA pathogenesis via indirect suppression of Tregs via the lower relative abundance of *B. fragilis* in these patients, but might equally be explained by the inflammatory milieu hypothesis. Genotyping of new onset RA patients showed that this increase was associated with HLA DRB1 4 shared epitope (SE) genotype [26]. This work replicated previous findings which - whilst not directly linked to RA, have linked host genetics with *Prevotella* and other taxa - *Pasteurellaceae*, and *Leptotrichia*, which were associated with SNPs encoding for ATP-binding cassettes, protein synthesis, cell division, and tumour suppression [40]. Similar association of *Prevotella* with HLA has been shown in mice [41]. These findings warrant further investigation. We know that the host genome impacts the microbiome: if microbiome changes are also mediated by genotype we could speculate that microbiome alteration appears before clinical disease manifests, and perhaps lies on the causal pathway to RA.

Zhang and collaborators undertook shotgun metagenomic sequencing on 212 faecal samples from 3 groups: 77 untreated RA patients matched with 80 unrelated healthy controls, 17 untreated RA patients matched with 17 healthy related controls, and a third group of 21 DMARD-treated RA patients [18]. To delineate features of the RA-associated gut microbiome, they identified 117,219 gene markers that were differentially enriched in RA patients versus controls and clustered the genes into metagenomic linkage groups (MLGs) on the basis of their correlated abundance variation among samples. They identified 88 MLGs that contained at least 100 genes, which separated RA-enriched and control-enriched MLGs. Of the MLGs comprising greater than 100 genes, RA was associated with MLGs containing *Lactobacillus salivaris, Clostridium asparagiforme, Gordonibacter pamelaeae, Eggerthella lenta, Lachnospiraceae bacterium, Bifidobacterium dentium, Lactobacillus spp* and *Ruminococcus lactaris*. The control group showed a negative correlation with *Haemophilus spp., K. pneumoniae* and *Bacteroides spp*., *B.bifidum* and *R.lactaris* [18], suggesting an antagonistic or mutually exclusive relationship and highlighting the interdependency of species especially when measured as relative abundances [42,43]. None of the MLGs containing 100 or more genes contained *Prevotella*, however when the authors compared the MLGs with the NCBI *Prevotella* reference genome, there was a trend towards increased abundance of *P.copri* as a function of RA duration in the first year of disease onset [18]. *Prevotella* may therefore be a particularly variable taxon, where bacterial genes and associated functions vary to a greater extent than in other taxa.

Zhang and collaborators also showed that treatment with RA therapeutics – disease modifying antirheumatic drugs (DMARDs), in which the inflammatory response is abated, was associated with a "normalised" more diverse microbiome; this is in contrast to the general observation that medication intake associates with reduced diversity [22]. This effect was observed for Methotrexate, as the most widely prescribed DMARD. Other drugs used by the cohort – Leflunomide,

Table 2
Studies of the microbiome in RA patients.

Study	Site	Cases	Controls	Methods
Shinenbaum et al., 1987 [117]	Gut	25 RA	25 healthy	Bacterial culture
Eerola et al., 1994 [118]	Gut	74 RA	91 healthy	Gas liquid chromatography analysis of bacterial cell fatty acid.
Toivanen et al., 2002 [44]	Gut	25 early RA	23 neuroinflammatory pain	16SrRNA oligonucleotide probes
Vaahtuvuo et al., 2008 [119]	Gut	50 RA	50 Fibromyalgia	Flow cytometry, 16s rRNA hybridisation, DNA staining
Scher et al., 2012 [5]	Oral	31 early RA 34 treated RA	18 healthy	16s
Scher et al., 2013 [26]	Gut	44 early RA 26 treated RA 16 psoriatic arthritis	28 healthy	16s
Liu et al., 2013 [120]	Gut	15 RA	15 controls	Quantitative RT-PCR
Zhang et al., 2015 [18]	Gut/Oral	94 early RA 21 treated RA	97 healthy	Metagenomes
Scher et al., 2016 [35]	Lung	20 RA 20 Sarcoidosis	28 Healthy	16s

Prednisolone, Hydroxychloroquine and Etanercept were not prescribed to enough participants to power analysis. Further study of specific DMARDS and at multiple time-points is needed. However, this supports the hypothesis that microbiome changes in RA are driven at least in part by systemic inflammation.

The above studies accounted for body mass index (BMI) and sex, which have known associations with gut microbiota and these factors showed no association with microbiota profile. The reported associations thus far are therefore unlikely to be driven by diet or obesity. However an influence of diet on RA via the microbiome is possible, and warrants further study. Similarly, whilst gender is controlled for, the independent role of gender in the link between the microbiome and RA has not been investigated. However an influence is feasible, via genetic differences (some associated SNPs lie on the sex chromosomes), diet and hormone fluctuation.

Prevotella copri is therefore the key candidate taxon in RA, with a finding of increased abundance in RA patients compared to controls replicated across 4 of 7 published epidemiological studies of the microbiome using next generation sequencing [5,18,26,44]. Moreover, *P.copri* was shown to mediate immune activation with 24% of RA patients having IgA or IgG responses to Pc-p27, an HLA-DR presented *P.copri* peptide, compared to 2% in healthy controls [37]. Further investigating the specificity of *P.copri* in immune activation in RA, *P.gingivalis* (shown previously to increase in RA), *Bacteroides fragilis* (shown previously to decrease in RA) and Escherechia coli (not previously implicated in RA), were shown to have no immune memory. While suggestive, the associations in human RA patients to date has been unable to elucidate whether the *Prevotella*-RA association is driven by host inflammation, or is in some part causal.

Mouse studies demonstrated that the *Prevotellecae* enriched microbiome of RA patients, when transferred to arthritis susceptible SKG mice, increases sensitivity to arthritis via activation of autoreactive T cells in the intestine. T cell differentiation and an increase in autoreactive T cells is a known driver of RA pathogenesis. Faeces from 3 RA patients and 3 healthy controls were analysed for microbiome composition, before transfer to SKG mice. After 20 weeks of colonisation, the total number of $CD4^+$ T cells and the number of pro-inflammatory IL-17 producing $CD4^+$ T cells in the large intestine were increased in RA-SKG mice compared with HC-SKG mice [45]. It remains unclear whether *P. Copri* is robustly associated with or actually causative in RA [16,26,36,37,39,46]. One explanation for the conflicting findings is that *Prevotella* is a particularly variable taxon, with different genes and biological characteristics even at the strain level, leading to different results in relation to *P.copri* in RA. The jury is therefore still out on the relevance of prevotella to RA pathogenesis, though current findings are promising. Associations with other taxa show weaker evidence, with lower reproducibility across studies. Overall, taxa associations, whilst useful, can only go some way to indicating the genomic functional capacity of the microbiome. Further work, especially incorporating metagenomics analysis of genetic capability of a microbiome will be critical to provide clarity.

1.2.1. CARD 9

The *caspase recruitment domain family member 9* gene (CARD9) which encodes an adaptor protein that integrates signals downstream of pattern recognition receptors [50], has been shown to be important in mouse models of arthritis [51]. Using two CARD9 knockout mouse models, with neutrophil-specific deletion, and wild-type control mice, the effect of CARD9 deficiency was investigated using the KBXN serum transfer model. Clinical signs of arthritis were quantified in CARD9 knockout mice which showed significant ankle thickening ($P = 0.0047$) and reduced grip strength ($P = 2 \times 10^{-4}$), compared to wild-type mice [51]. To date, no GWAS has shown CARD9 to be a risk allele for RA in humans, but it is an interesting model gene which has relevance in other autoimmune diseases including inflammatory bowel disease (IBD) [52], ankylosing spondylitis [53], IgA nephropathy [50] and colitis [54].

CARD9 knockout mice exhibited downregulation in IL-22 signalling, resulting in impaired recovery from colitis when compared to wild-type mice. Analysis of the gut microbiota revealed no difference in beta diversity, but highlighted a decrease in abundance of *Adlercreutzia* (genus), *Actinobacteria* (phylum), and *Lactobacillus reuteri* (species), indicating an in influence of CARD9 genotype on the gut microbiome in these mice. Transfer of the CARD9 knockout mouse microbiota to germ-free wild type mice resulted in an exacerbation of colitis, to a similar degree as CARD9 knock out mice – suggesting a causative, rather than correlative, relationship between the microbiome and IL-22 mediated inflammation.

The gut microbiota of CARD9 knockout mice showed impaired catabolism of tryptophan, a downstream effect of which is IL-22 production, and this is suggested to be a key underlying mechanism [54]. IL-22 is a Th17 cytokine, which in addition to an integral role in maintaining the gut barrier, has diverse functions in balance with pro-inflammatory IL-17, which vary according to tissue type and duration of expression. IL-22 promotes wound healing and tissue homeostasis acutely, however chronic unabated expression is associated with a number of inflammatory conditions. The role of neutrophils in RA has been highlighted in a recent review [55].

1.3. The oral microbiome in RA

1.3.1. Shared epitope interaction with the oral microbiome

One of the earliest links between commensal microbiota and RA pathogenesis was shown in the oral microbiome. Individuals with RA had a higher incidence of periodontal disease – linked to oral microbiome dysbiosis, and that periodontal disease treatment improved RA

symptoms [5,47,48]. This is highly plausible because of the known interaction between the oral microbiome, co-occurring with periodontal disease and progression to clinical RA, secondary to shared epitope HLA-DRB1 genotype. The oral and lung mucosa have been proposed to be the primary sites of protein citrullination in RA, via oral microbiome changes [5] and smoking status [4], respectively. The host immune response to citrullinated proteins is mediated by the shared epitope, which encodes the binding motif of MHC II. Variation in MHCII results in altered immune response to extracellular antigens, and carriers of the shared epitope have enhanced response to citrullinated proteins, and subsequent increase in ACPA. RA patients have been shown to have a 5.7 fold increased risk of periodontal disease (95% CI 2.35–13.84) in a stepwise logistic regression, including other predictors of periodontal disease – age, education, smoking, alcohol consumption level and BMI, only age and RA remained as significant predictors [48]. The oral microbiome may therefore be the primary mediator of protein citrullination, having greater influence even than smoking.

Bacterial taxa associated with RA may provide the mechanistic link for this association. *Porphyromonas gingivalis, is* an oral commensal found in increased in abundance in RA patients [5,18], and it is active in citrullination of host proteins – providing a precursor step for production of specific antibodies [18], and mediating synovial inflammation (see Fig. 2) [46]. Further, *Porphyromonas gingivalis* has evolved to alter its microenvironment within the oral microbiome by modulating the host TLR2 pathway to uncouple bacterial clearance from inflammation and therefore exacerbate the microbial ecological niche [49].

1.5. General mechanisms linking the microbiome and RA

The accumulating evidence implicating the microbiome in RA pathogenesis has prompted investigation into the underlying mechanisms, of which a number have been proposed: molecular mimicry; outer membrane vesicles (OMVs); T cell differentiation; epigenetic modification; immune priming (see Fig. 3), and a role in immune ageing (see Box 1).

The gut microbiome produces a variety of metabolites, including small organic acids, bile acids, vitamins, choline metabolites, and lipids [64,65]. The plethora of small molecules produced, alongside microbial cellular components, share some structural similarity with the host. Such molecular mimicry refers to the similarity of bacterial peptides of RA associated antigens, or to affinity of bacterial peptides to host receptors [18]. Molecules associated with bacterial cell to cell communication – quorum sensing, may also influence cellular processes within the host (see Box 1).

For example, *P.gingivalis,* which is found in increased abundance in the oral microbiome in RA (and also found in the gut) shares 82% homology of α-enolase with human α-enolase at the immunodominant region. Human antibodies against bacterial elonase also recognise human α-enolase, promoting further antibody production [66]. In RA patients, levels of anti-citrullinated human α-enolase antibodies and bacterial α-enolase are shown to correlate with one another ($R^2 = 0.0803$, $P < 0.0001$) [67].

Molecular mimicry also promotes autoreactive T cell activation and proliferation. *E. coli* heat shock protein DnaJ contains a QKRAA motif that is also present in the HLA-DRB1 shared epitopes [66]. DnaJ strongly activated RA synovial T cells which had passed the positive selection in the thymus through weak binding with the corresponding HLA epitopes [68].

OMVs produced by gram negative bacteria modify the local environment to facilitate bacterial proliferation, signal between bacterial species and have been shown to communicate directly with host cells [69–71]. In the host, OMVs affect intracellular signalling [71] and overall metabolic profile [72]. Thus, investigation of OMVs may be an important step in understanding the link between the microbiome and host. Interestingly, OMVs of pathogenic and non-pathogenic strains of the same species manifest differing metabolic associations [73].

The anti-inflammatory influence of the microbiome may also play a role in RA. Dietary poly- and oligosaccharides resistant to upper gut digestion pass to the distal gut where they serve as a source of carbon and energy for gut bacteria. Through fermentative reactions, the gut microbiota can metabolize complex carbohydrates to produce small organic acids, the majority of which are comprised of the short-chain fatty acids (SCFAs) -acetate, propionate, and butyrate.

SCFAs, and butyrate in particular may influence host physiology, as these metabolites are linked to expansion of Tregs, and a protective anti-inflammatory role is proposed for them.

1.5. Considerations for future studies

1.5.1. Limitations of mouse models in microbiome research

Mouse models are widely used in microbiome research and can be informative – particularly when used to understand or replicate a specific mechanism. However, there are key differences between mouse and human microbiome physiology which are seldom discussed. The mouse and human genome are separated by more than 90 million years of evolution, during which there has been substantial change in the immune system and its regulation. The GI tract anatomy and physiology

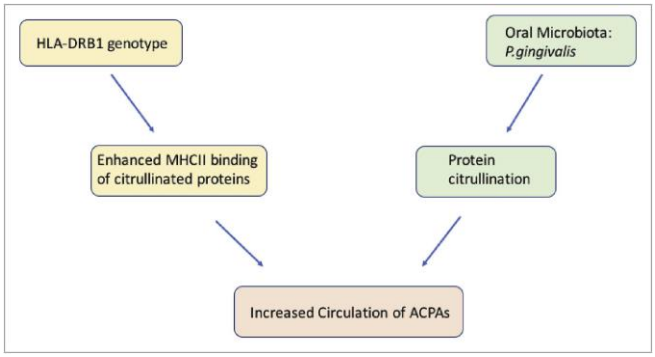


Fig. 2. Interaction between host genetics and the oral microbiome in RA. Mechanism by which HLA-DRB1 genotype and *Porphyromonas gingivalis* in the oral microbiota can interact to contribute to RA pathology, via upregulation of circulating ACPA.

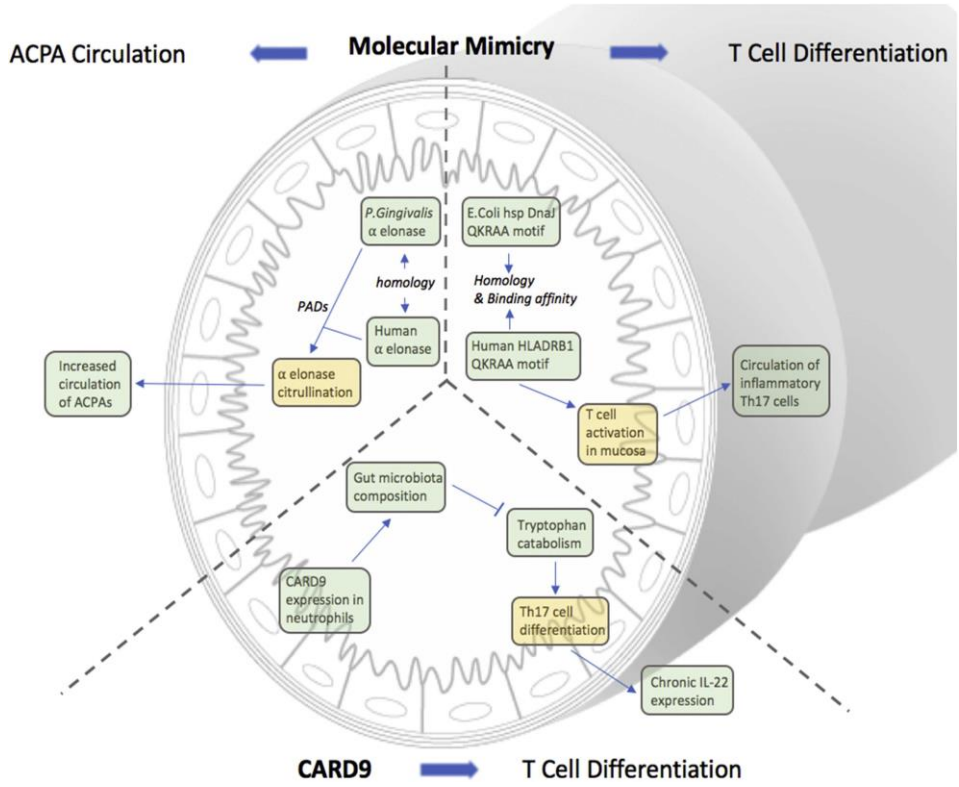


Fig. 3. Summary of mechanisms by which molecular mimicry of the gut microbiota, and CARD9 genotype can contribute to systemic inflammation.

Box 1
RA as a Model of Ageing **–** Potential role of epigenetics and the microbiome

Patients with RA prematurely show a number of features of an aged immune system, and RA has therefore been proposed as a model of immune ageing. There are a wide range of disruptions to the delicate balance within the immune system which occur with ageing, mediated in part by epigenetic changes, and having potential impact on the host response to the microbiome [56]. Key changes are compromising of epithelial layer integrity through disruption of tight junctions [57], and immune-senescence [56,58], which may be secondary to a reduction in telomerase [56,59]. Loss of function of telomerase in lymphocytes, leads to loss of the unique ability of these cells to elongate telomeres, a consequence of which is cellular senescence [60]. A second pathway, independent of telomerase, is senescence secondary to genomic instability and prolonged activation of the DNA damage response (DDR). The DDR pathway has been shown to be particularly important in peripheral blood mononuclear cells (PBMC) and both naïve and memory $CD4^+$ T cells, but not neutrophils, in RA patients [61]. Cellular senescence of T cells results in clonal expansion of sets of naïve T cells. Early in the onset of RA there is clonal expansion of both $CD4^+$ T and C8+ T cells. This is accompanied by a relative lack of expression of co-stimulatory molecules from memory T cells, particularly CD28, which is required for efficient T-cell activation and proliferation. A lower level of CD28 is associated with a pro-inflammatory phenotype, increased cytotoxicity and increased rates of tissue infiltration [56,61]. Other immune changes in both ageing and RA include altered patterns of DNA methylation and therefore gene expression, chromatin remodelling, failure of protein homeostasis, altered nutrient sensing and mitochondrial dysfunction [56]. There are therefore diverse differences, having far reaching physiological effects. Animal models have been used to determine which intracellular pathways are most implicated in ageing, and a key identified difference is defective transgenic growth factor β (TGFβ) signalling. More specifically, downstream, SMAD3 signalling has been shown to promote cartilage damage [62]. The microbiome may contribute to these shifts in immune phenotype; expansion of the pro-inflammatory cytokine milieu in the host is suggested as the primary mechanism of immune ageing [63].

of the mouse is markedly different to that of humans, and in particular the presence of the glandular forestomach in the mouse, with its biofilm of *lactobacillus spp,* and mucus trap where mucus and bacteria are recycled to the cecum. There are differences in morphology and retention time, and mice engage in coprophagy, and transfer gut microbiota between each other when housed together. These factors confer differences in mouse vs human microbiome, and in physiological response to bacteria. The presence of taxonomic differences in the murine versus human microbiome may have resulted in over-interpretation of the clinical relevance of findings shown in mouse models. There are numerous references to the importance of *segmented filamentous bacteria* (SFB) within the RA microbiome literature, since it was shown that introduction of SFB in a mouse model regulated TH17 differentiation [74]. This is often cited as convincing evidence of the microbiome inducing T cell differentiation. However, SFB, also known as *Candidatus arthromitus* within the Greengenes database, are usually only present in humans during early life, and so the relevance in human RA is at best unclear, and whether other species have comparable effect in humans remains open to debate.

In murine models of inflammatory arthritis germ free mice fail to develop the diseased state [45], in contrast to microbiota exposed arthritis model mice which ubiquitously develop symptoms, suggesting an integral role in immune development of commensal microbiota [75]. Germ free mice exhibit immune differences, culminating in a much-dampened immune response. Therefore, it is unsurprising that these mice do not develop inflammatory or autoimmune disease. The relative disease progression of mice gavaged with RA-associated microbiota compared to germ free mice should therefore be interpreted with caution and comparison to mice with a different ("healthy") humanised microbiome might be preferable (see section 1.5.2).

Large-scale human observational studies in RA are underway, including the use of family- and twin studies to unpick the contributions of host genetics and microbiome. In addition, population based studies linked to health records are now incorporating microbiome assessment, which will allow the interrogation of the microbiome prior to diagnosis retrospectively. Studies in at risk groups, such as those at genetic risk, those with periodontal disease and smokers may provide insight as to the temporal relationship between microbiome alterations and altered inflammation. Low-risk interventional studies in humans are around the corner and will be needed to ascertain whether effects are relevant to the development of RA or bystanders of an altered milieu.

1.5.2. Developing bioinformatic approaches for microbiome data

Next generation sequencing has provided the means to access a wealth of information relating to commensal bacterial communities, however there are a wide variety of technological and analytical methods available – it is important to understand the benefits and limitations of the most frequently used methods: operational taxonomic units (OTUs), amplicon sequence variants (ASVs) and metagenomics. The most frequently used approach in microbiome sequencing is16s. This refers to sequencing of bacterial ribosomal RNA (rRNA) gene, using methods which take advantage of the particular structure of the 16s ribosomal component [76]. 16s rRNA is a component of 30s small subunit of prokaryotic ribosome, of which there are 9 variable regions – varying between phylogenies, which are each flanked by highly conserved regions. DNA is first extracted from the biological sample. Following this, primers linked to an identifier barcode, specific to highly conserved binding sites, provide a template for PCR amplification. An alternative approach is whole genome metagenomic sequencing, this is more costly but provides important additional information - particularly bacterial genes present [77]. Classically, following sequencing data are assigned to operational taxonomic units (OTUs), in which sequences are binned (grouped) together according to a similarity threshold. There are a number of alternative methods picking OTUs – open reference, closed reference or de novo approaches can be taken. These approaches differ in the use of a database (e.g Greengenes or SILVA) as a reference when performing clustering. Within these approaches, a number of algorithms may be used. The de-novo approach, whilst computationally less efficient (requiring pair-wise comparisons between sequence reads), has been shown to produce OTUs that are more representative of functional microbial units [78]. The OTUs in conjunction with a genome reference database are then used to assign taxonomy, determining which bacteria are present in the samples. Amplicon Sequence Variants (ASVs), also known as exact sequence variants, offer an alternative approach to OTUs, and hold a number of advantages over traditional OTU methods [79]. Briefly, ASVs are generated by using the error rate within the dataset to infer true biological sequences, and group identical sequences exactly, rather than to a similarity threshold. There are a number of advantages to this approach, demonstrated by increased sensitivity and specificity of ASVs compared to the most often used methods of OTU generation [79,80]. ASVs overcome other key limitations of OTUs and allow for valid direct comparability between datasets [80], and accurate taxonomic assignment at the species level [79]. The direct comparability between datasets, in conjunction with correcting for 'batch effects' on analysis could allow for the merging of multiple RA microbiome datasets, which could shed light on the conflicting taxa associations reported. 16s analysis is inferior however to metagenomic sequencing, which although at greater expense, provides key additional information – genes present and full species assignment, allowing for much richer functional analysis. Moving forward, ASVs show exciting potential in 16s analysis, and could advance understanding of the role of the microbiome in RA, both through analysis of existing datasets and through use in future studies.

1.6. Sample collection, storage and processing

In most studies, the means of measurement of the gut microbiome is the microbial composition of stool samples. Although the faecal microbiome provides a useful indicator of the gut microbiome, they are of course distinct entities with differences starting in the mucosal layer, epithelium, lumen in small intestine through to stool. In studies of the gut microbiome therefore, a useful additional measure may be the use of colonic biopsy. Differences in sample storage, and particularly the immediacy of freezing sample, temperature on freezing or use of an RNA inhibitor with sample add to the difficulty of comparing results across studies. This is difficult to address, but newer bioinformatic approaches as described above (see section 1.5.2) may go some way towards a solution.

1.7. Therapeutic modulation of the microbiome in RA

Current understanding of the microbiome in RA does not feasibly allow development of therapeutics at this stage. Moving forward, there is potential in the future for the microbiome to be useful as a target in RA, either as a target for modulation, or as a biomarker of potential for disease progression in arthralgia. Modulation may be possible via probiotics comprised of bacteria with beneficial functions, or precision editing, for example through use of bacteriophages. Given that RA seems to associate more with abundance of pathogenic bacteria, as opposed to a deficit in beneficial bacteria, precision editing of the microbiome may be a more likely option, for example through use of bacteriophages.

Current understanding of the RA microbiome is at an early stage, and as would be expected the pre-emptive trials of general use, non-RA specific probiotics have been inconclusive [12]. Use of the microbiome as a biomarker, to detect those at greater risk of progression from arthralgia to RA, is also feasible – if microbiome changes are present before onset of disease, and would allow for earlier intervention in these patients to improve clinical outcome [3].

2. Conclusions

There are compelling associations between the microbiome and RA, although the current evidence is far from conclusive that the microbiome causes RA. Strategic future studies replicating previous findings and addressing the gaps in the current knowledge are required. In particular it will be important to determine the influence of disease modifying RA medication on the microbiome. Host genetics may provide the link between the microbiome and RA and is a particular challenge to address, although current findings are suggestive of an important influence which may be mediated by the host immune system which could be ameliorated. A better understanding of whether associations described thus far are confounded by host genetics will shed further light on the role of the microbiome in rheumatoid arthritis.

Funding

Twins UK receives funding from the Wellcome Trust European Community's Seventh Framework Programme (FP7/2007-2013 to TwinsUK); the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. Cristina Menni is funded by the MRC AimHy (grant MR/M016560/1), Frances M.K Williams and Philippa M. Wells are funded by Arthritis Research UK Special Strategic Award (grant 21227), Claire J. Steves is funded by the Wellcome Trust (grant WT081878MA).

Acknowledgements

The authors of this paper wish to express our appreciation to all study participants of the TwinsUK cohort. We thank ML Matey-Hernandez for generating the circos plot.

References

[1] A. Gibofsky, Epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis: a Synopsis, Am. J. Manag. Care 20 (2014) S128–S135.

[2] T.D. Wilsdon, C.L. Hill, Managing the drug treatment of rheumatoid arthritis, Aust. Prescr. 40 (2017) 51–58, https://doi.org/10.18773/austprescr.2017.012.

[3] HW van Steenbergen, JAP da Silva, T.W.J. Huizinga, H. Mil AHM van der, Preventing progression from arthralgia to arthritis: targeting the right patients, Nat. Rev. Rheumatol. (2017) 185, https://doi.org/10.1038/nrrheum.2017.185 nrrheum.2017.

[4] N.J. Bernard, Synergy of environmental and genetic risk in RA, Nat. Rev. Rheumatol. 14 (2018) 319, https://doi.org/10.1038/s41584-018-0015-7.

[5] J.U. Scher, C. Ubeda, M. Equinda, R. Khanin, Y. Buischi, A. Viale, et al., Periodontal disease and the oral microbiota in new-onset rheumatoid arthritis, Arthritis Rheum. 64 (2012) 3083–3094, https://doi.org/10.1002/art.34539.

[6] I.B. McInnes, G. Schett, The pathogenesis of rheumatoid arthritis, N. Engl. J. Med. 365 (2011) 2205–2219, https://doi.org/10.1056/NEJMra1004965.

[7] G.S. Firestein, I.B. McInnes, Immunopathogenesis of rheumatoid arthritis, Immunity 46 (2017) 183–196, https://doi.org/10.1016/j.immuni.2017.02.006.

[8] K. Kim, S.-Y. Bang, H.-S. Lee, S.-C. Bae, Update on the genetic architecture of rheumatoid arthritis, Nat. Rev. Rheumatol. 13 (2017) 13–24, https://doi.org/10.1038/nrrheum.2016.176.

[9] A.J. Svendsen, N.V. Holm, K. Kyvik, P.H. Petersen, P. Junker, Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population, BMJ 324 (2002) 264–266.

[10] A.P. Webster, D. Plant, S. Ecker, F. Zufferey, J.T. Bell, A. Feber, et al., Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins, Genome Med. 10 (2018) 64, https://doi.org/10.1186/s13073-018-0575-9.

[11] C. Wilson, H. Tiwana, A. Ebringer, Molecular mimicry between HLA-DR alleles associated with rheumatoid arthritis and Proteus mirabilis as the Aetiological basis for autoimmunity, Microb. Infect. 2 (2000) 1489–1496.

[12] G. Horta-Baas, M.D.S. Romero-Figueroa, A.J. Montiel-Jarquín, M.L. Pizano-Zárate, J. García-Mena, N. Ramírez-Durán, Intestinal dysbiosis and rheumatoid arthritis: a link between gut microbiota and the pathogenesis of rheumatoid arthritis, J Immunol Res 2017 (2017) 4835189, https://doi.org/10.1155/2017/4835189.

[13] R.K. Mandal, T. Jiang, A.A. Al-Rubaye, D.D. Rhoads, R.F. Wideman, J. Zhao, et al., An investigation into blood microbiota and its potential association with Bacterial Chondronecrosis with Osteomyelitis (BCO) in Broilers, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep25882.

[14] T.J. Hieken, J. Chen, T.L. Hoskin, M. Walther-Antonio, S. Johnson, S. Ramaker, et al., The microbiome of aseptically collected human breast tissue in benign and malignant disease, Sci. Rep. 6 (2016) 30751, https://doi.org/10.1038/srep30751.

[15] C.A. Thaiss, N. Zmora, M. Levy, E. Elinav, The microbiome and innate immunity, Nature 535 (2016) 65–74, https://doi.org/10.1038/nature18847.

[16] J.T. Rosenbaum, M.J. Asquith, The microbiome: a revolution in treatment for rheumatic diseases? Curr. Rheumatol. Rep. 18 (2016) 62, https://doi.org/10.1007/s11926-016-0614-8.

[17] L. Lin, J. Zhang, Role of intestinal microbiota and metabolites on gut homeostasis and human diseases, BMC Immunol. 18 (2017) 2, https://doi.org/10.1186/s12865-016-0187-3.

[18] X. Zhang, D. Zhang, H. Jia, Q. Feng, D. Wang, D. Liang, et al., The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment, Nat. Med. 21 (2015) 895–905, https://doi.org/10.1038/nm.3914.

[19] L.A. David, C.F. Maurice, R.N. Carmody, D.B. Gootenberg, J.E. Button, B.E. Wolfe, et al., Diet rapidly and reproducibly alters the human gut microbiome, Nature 505 (2014) 559–563, https://doi.org/10.1038/nature12820.

[20] J.K. Goodrich, E.R. Davenport, M. Beaumont, M.A. Jackson, R. Knight, C. Ober, et al., Genetic determinants of the gut microbiome in UK twins, Cell Host Microbe 19 (2016) 731–743, https://doi.org/10.1016/j.chom.2016.04.017.

[21] C.J. Steves, S. Bird, F.M.K. Williams, T.D. Spector, The microbiome and musculoskeletal conditions of aging: a review of evidence for impact and potential therapeutics, J Bone Miner Res Off J Am Soc Bone Miner Res 31 (2016) 261–269, https://doi.org/10.1002/jbmr.2765.

[22] M.A. Jackson, S. Verdi, M.-E. Maxan, C.M. Shin, J. Zierer, R.C.E. Bowyer, et al., Gut microbiota associations with common diseases and prescription medications in a population-based cohort, Nat. Commun. 9 (2018) 2655, https://doi.org/10.1038/s41467-018-05184-7.

[23] W. Zhu, M.G. Winter, M.X. Byndloss, L. Spiga, B.A. Duerkop, E.R. Hughes, et al., Precision editing of the gut microbiota ameliorates colitis, Nature 553 (2018) 208–211, https://doi.org/10.1038/nature25172.

[24] B. Paul, S. Barnes, W. Demark-Wahnefried, C. Morrow, C. Salvador, C. Skibola, et al., Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases, Clin. Epigenet. 7 (2015), https://doi.org/10.1186/s13148-015-0144-7.

[25] M.C. Dao, A. Everard, J. Aron-Wisnewsky, N. Sokolovska, E. Prifti, E.O. Verger, et al., Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology, Gut (2015), https://doi.org/10.1136/gutjnl-2014-308778 gutjnl-2014-308778.

[26] J.U. Scher, A. Sczesnak, R.S. Longman, N. Segata, C. Ubeda, C. Bielski, et al., Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis, ELife 2 (2013) e01202, https://doi.org/10.7554/eLife.01202.

[27] C. Duvallet, S.M. Gibbons, T. Gurry, R.A. Irizarry, E.J. Alm, Meta-analysis of gut microbiome studies identifies disease-specific and shared responses, Nat. Commun. 8 (2017) 1784, https://doi.org/10.1038/s41467-017-01973-8.

[28] C. Menni, J. Zierer, A.M. Valdes, T.D. Spector, Mixing omics: combining genetics and metabolomics to study rheumatic diseases, Nat. Rev. Rheumatol. (2017), https://doi.org/10.1038/nrrheum.2017.5 advance online publication.

[29] Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery, Nature 506 (2014) 376–381, https://doi.org/10.1038/nature12873.

[30] G.L. Burn, L. Svensson, C. Sanchez-Blanco, M. Saini, A.P. Cope, Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? FEBS Lett. 585 (2011) 3689–3698, https://doi.org/10.1016/j.febslet.2011.04.032.

[31] X.-K. Yang, J. Liu, J. Liu, Y. Liang, W.-D. Xu, R.-X. Leng, et al., Associations between PADI4 gene polymorphisms and rheumatoid arthritis: an updated meta-analysis, Arch. Med. Res. 46 (2015) 317–325, https://doi.org/10.1016/j.arcmed.2015.05.011.

[32] E.F. Remmers, R.M. Plenge, A.T. Lee, R.R. Graham, G. Hom, T.W. Behrens, et al., STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus, N. Engl. J. Med. 357 (2007) 977–986, https://doi.org/10.1056/NEJMoa073003.

[33] Y.M. Hussein, R.H. Mohamed, H.F. Pasha, E.E. El-Shahawy, S.S. Alzahrani, Association of tumor necrosis factor alpha and its receptor polymorphisms with rheumatoid arthritis in female patients, Cell. Immunol. 271 (2011) 192–196, https://doi.org/10.1016/j.cellimm.2011.06.023.

[34] A.M. Walsh, J.W. Whitaker, C.C. Huang, Y. Cherkas, S.L. Lamberth, C. Brodmerkel, et al., Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations, Genome Biol. 17 (2016) 79, https://doi.org/10.1186/s13059-016-0948-6.

[35] J.U. Scher, V. Joshua, A. Artacho, S. Abdollahi-Roodsaz, J. Öckinger, S. Kullberg, et al., The lung microbiota in early rheumatoid arthritis and autoimmunity, Microbiome 4 (2016) 60, https://doi.org/10.1186/s40168-016-0206-x.

[36] H.K. Pedersen, V. Gudmundsdottir, H.B. Nielsen, T. Hyotylainen, T. Nielsen, B.A.H. Jensen, et al., Human gut microbes impact host serum metabolome and insulin sensitivity, Nature 535 (2016) 376–381, https://doi.org/10.1038/nature18646.

[37] A. Pianta, S. Arvikar, K. Strle, E.E. Drouin, Q. Wang, C.E. Costello, et al., Evidence for immune relevance of prevotella copri, a gut microbe, Patients with Rheumatoid Arthritis. Arthritis Rheumatol Hoboken NJ, 2016, https://doi.org/10.1002/art.40003.

[38] D. Kim, W. Kim, Editorial: can prevotella copri Be a causative pathobiont in rheumatoid arthritis? Arthritis Rheumatol Hoboken Nj 68 (2016) 2565–2567, https://doi.org/10.1002/art.39807.

[39] Hofer U. Microbiome, Pro-inflammatory prevotella? Nat. Rev. Microbiol. 12 (2014), https://doi.org/10.1038/nrmicro3180 5–5.

[40] R. Blekhman, J.K. Goodrich, K. Huang, Q. Sun, R. Bukowski, J.T. Bell, et al., Host genetic variation impacts microbiome composition across human body sites, Genome Biol. 16 (2015) 191, https://doi.org/10.1186/s13059-015-0759-1.

[41] A. Gomez, D. Luckey, C.J. Yeoman, E.V. Marietta, M.E.B. Miller, J.A. Murray, et al., Loss of sex and age driven differences in the gut microbiome characterize arthritis-susceptible *0401 mice but not arthritis-resistant *0402 mice, PLoS One 7 (2012) e36095, https://doi.org/10.1371/journal.pone.0036095.

[42] J. Friedman, L.M. Higgins, J. Gore, Community structure follows simple assembly rules in microbial microcosms, Nat Ecol Evol 1 (2017) 0109, https://doi.org/10.1038/s41559-017-0109.

[43] M.K.A. Gavina, T. Tahara, K. Tainaka, H. Ito, S. Morita, G. Ichinose, et al., Multi-species coexistence in Lotka-Volterra competitive systems with crowding effects, Sci. Rep. 8 (2018) 1198, https://doi.org/10.1038/s41598-017-19044-9.

[44] P. Toivanen, S. Vartiainen, J. Jalava, Intestinal anaerobic bacteria in early rheumatoid arthritis, RA. Arthritis Res 4 (2002) 5.

[45] Y. Maeda, T. Kurakawa, E. Umemoto, D. Motooka, Y. Ito, K. Gotoh, et al., Dysbiosis contributes to arthritis development via activation of autoreactive T cells in the intestine, Arthritis Rheumatol 68 (2016) 2646–2661, https://doi.org/10.1002/art.39783.

[46] D. Kim, W. Kim, Editorial: can prevotella copri Be a causative pathobiont in rheumatoid arthritis? Arthritis Rheumatol Hoboken Nj 68 (2016) 2565–2567, https://doi.org/10.1002/art.39807.

[47] P. Ortiz, N.F. Bissada, L. Palomo, Y.W. Han, M.S. Al-Zahrani, A. Panneerselvam, et al., Periodontal therapy reduces the severity of active rheumatoid arthritis in patients treated with or without tumor necrosis factor inhibitors, J. Periodontol. 80 (2009) 535–540, https://doi.org/10.1902/jop.2009.080447.

[48] N. Pischon, T. Pischon, J. Kröger, E. Gülmez, B.-M. Kleber, J.-P. Bernimoulin, et al., Association among rheumatoid arthritis, oral hygiene, and periodontitis, J. Periodontol. 79 (2008) 979–986, https://doi.org/10.1902/jop.2008.070501.

[49] T. Maekawa, J.L. Krauss, T. Abe, R. Jotwani, M. Triantafilou, K. Triantafilou, et al., Porphyromonas gingivalis manipulates complement and TLR signaling to uncouple bacterial clearance from inflammation and promote dysbiosis, Cell Host Microbe 15 (2014) 768–778, https://doi.org/10.1016/j.chom.2014.05.012.

[50] K. Kiryluk, Y. Li, F. Scolari, S. Sanna-Cherchi, M. Choi, M. Verbitsky, et al., Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens, Nat. Genet. 46 (2014) 1187–1196, https://doi.org/10.1038/ng.3118.

[51] T. Németh, K. Futosi, C. Sitaru, J. Ruland, A. Mócsai, Neutrophil-specific deletion of the CARD9 gene expression regulator suppresses autoantibody-induced inflammation *in vivo*, Nat. Commun. 7 (2016) 11004, https://doi.org/10.1038/ncomms11004.

[52] M.A. Rivas, M. Beaudoin, A. Gardet, C. Stevens, Y. Sharma, C.K. Zhang, et al., Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease, Nat. Genet. 43 (2011) 1066–1073, https://doi.org/10.1038/ng.952.

[53] J.J. Pointon, D. Harvey, T. Karaderi, L.H. Appleton, C. Farrar, M.A. Stone, et al., Elucidating the chromosome 9 association with AS; CARD9 is a candidate gene, Genes Immun. 11 (2010) 490–496, https://doi.org/10.1038/gene.2010.17.

[54] B. Lamas, M.L. Richard, V. Leducq, H.-P. Pham, M.-L. Michel, G. Da Costa, et al., CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands, Nat. Med. 22 (2016) 598–605, https://doi.org/10.1038/nm.4102.

[55] I. Cecchi, I. Arias de la Rosa, E. Menegatti, D. Roccatello, E. Collantes-Estevez, C. Lopez-Pedrera, et al., Neutrophils: novel key players in Rheumatoid Arthritis. Current and future therapeutic targets, Autoimmun. Rev. (2018), https://doi.org/10.1016/j.autrev.2018.06.006.

[56] P. Chalan, A. van den Berg, B.-J. Kroesen, L. Brouwer, A. Boots, Rheumatoid arthritis, immunosenescence and the hallmarks of aging, Curr. Aging Sci. 8 (2015) 131–146.

[57] M.S. Lustgarten, Classifying aging as a disease: the role of microbes, Front. Genet. 7 (2016) 212, https://doi.org/10.3389/fgene.2016.00212.

[58] A. Pera, C. Campos, N. López, F. Hassouneh, C. Alonso, R. Tarazona, et al., Immunosenescence: implications for response to infection and vaccination in older people, Maturitas 82 (2015) 50–55, https://doi.org/10.1016/j.maturitas.2015.05.004.

[59] S.O. Schonland, C. Lopez, T. Widmann, J. Zimmer, E. Bryl, J.J. Goronzy, et al., Premature telomeric loss in rheumatoid arthritis is genetically determined and involves both myeloid and lymphoid cell lineages, Proc. Natl. Acad. Sci. 100 (2003) 13471–13476, https://doi.org/10.1073/pnas.2233561100.

[60] S.E. Steer, F.M.K. Williams, B. Kato, J.P. Gardner, P.J. Norman, M.A. Hall, et al., Reduced telomere length in rheumatoid arthritis is independent of disease activity and duration, Ann. Rheum. Dis. 66 (2007) 476–480, https://doi.org/10.1136/ard.2006.059188.

[61] P.J. Hohensinner, J.J. Goronzy, C.M. Weyand, Telomere dysfunction, autoimmunity and aging, Aging Dis 2 (2011) 524–537.

[62] P.M. van der Kraan, The changing role of TGFβ in healthy, ageing and osteoarthritic joints, Nat. Rev. Rheumatol. 13 (2017) 155–163, https://doi.org/10.1038/nrrheum.2016.219.

[63] C.M. Weyand, Z. Yang, J.J. Goronzy, T cell aging in rheumatoid arthritis, Curr. Opin. Rheumatol. 26 (2014) 93–100, https://doi.org/10.1097/BOR.0000000000000011.

[64] M.S. Donia, M.A. Fischbach, HUMAN MICROBIOTA. Small molecules from the human microbiota, Science 349 (2015) 1254766, https://doi.org/10.1126/science.1254766.

[65] P.C. Dorrestein, S.K. Mazmanian, R. Knight, Finding the missing links among metabolites, microbes, and the host, Immunity 40 (2014) 824–832, https://doi.org/10.1016/j.immuni.2014.05.015.

[66] S. Li, Y. Yu, Y. Yue, Z. Zhang, K. Su, Microbial infection and rheumatoid arthritis, J. Clin. Cell. Immunol. 4 (2013), https://doi.org/10.4172/2155-9899.1000174.

[67] K. Lundberg, A. Kinloch, B.A. Fisher, N. Wegner, R. Wait, P. Charles, et al., Antibodies to citrullinated alpha-enolase peptide 1 are specific for rheumatoid arthritis and cross-react with bacterial enolase, Arthritis Rheum. 58 (2008) 3009–3019, https://doi.org/10.1002/art.23936.

[68] S. Albani, E.C. Keystone, J.L. Nelson, W.E. Ollier, A. La Cava, A.C. Montemayor, et al., Positive selection in autoimmunity: abnormal immune responses to a bacterial dnaJ antigenic determinant in patients with early rheumatoid arthritis, Nat. Med. 1 (1995) 448–452.

[69] T.N. Ellis, M.J. Kuehn, Virulence and immunomodulatory roles of bacterial outer membrane vesicles, Microbiol Mol Biol Rev MMBR 74 (2010) 81–94, https://doi.org/10.1128/MMBR.00031-09.

[70] R. Stentz, N. Horn, K. Cross, L. Salt, C. Brearley, D.M. Livermore, et al., Cephalosporinases associated with outer membrane vesicles released by Bacteroides spp. protect gut pathogens and commensals against β-lactam antibiotics, J. Antimicrob. Chemother. 70 (2015) 701–709, https://doi.org/10.1093/jac/dku466.

[71] R. Stentz, S. Osborne, N. Horn, A.W.H. Li, I. Hautefort, R. Bongaerts, et al., A bacterial homolog of a eukaryotic inositol phosphate signaling enzyme mediates cross-kingdom dialog in the mammalian gut, Cell Rep. 6 (2014) 646–656, https://doi.org/10.1016/j.celrep.2014.01.021.

[72] W.A. Bryant, R. Stentz, G. Le Gall, M.J.E. Sternberg, S.R. Carding, T. Wilhelm, In silico analysis of the small molecule content of outer membrane vesicles produced by Bacteroides thetaiotaomicron indicates an extensive metabolic link between microbe and host, Front. Microbiol. 8 (2017), https://doi.org/10.3389/fmicb.2017.02440.

[73] N.B. Zakharzhevskaya, A.A. Vanyushkina, I.A. Altukhov, A.L. Shavarda, I.O. Butenko, D.V. Rakitina, et al., Outer membrane vesicles secreted by pathogenic and nonpathogenic Bacteroides fragilis represent different metabolic activities, Sci. Rep. 7 (2017) 5008, https://doi.org/10.1038/s41598-017-05264-6.

[74] H.-J. Wu, I.I. Ivanov, J. Darce, K. Hattori, T. Shima, Y. Umesaki, et al., Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells, Immunity 32 (2010) 815–827, https://doi.org/10.1016/j.immuni.2010.06.001.

[75] A. Paun, J.S. Danska, Immuno-ecology: how the microbiome regulates tolerance and autoimmunity, Curr. Opin. Immunol. 37 (2015) 34–39, https://doi.org/10.1016/j.coi.2015.09.004.

[76] B. Yang, Y. Wang, P.-Y. Qian, Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis, BMC Bioinf. 17 (2016) 135, https://doi.org/10.1186/s12859-016-0992-y.

[77] R. Ranjan, A. Rani, A. Metwally, H.S. McGee, D.L. Perkins, Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing, Biochem. Biophys. Res. Commun. 469 (2016) 967–977, https://doi.org/10.1016/j.bbrc.2015.12.083.

[78] M.A. Jackson, J.T. Bell, T.D. Spector, C.J. Steves, A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units, PeerJ 4 (2016) e2341, https://doi.org/10.7717/peerj.2341.

[79] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: high resolution sample inference from Illumina amplicon data, Nat. Methods 13 (2016) 581–583, https://doi.org/10.1038/nmeth.3869.

[80] B.J. Callahan, P.J. McMurdie, S.P. Holmes, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis, ISME J. 11 (2017) 2639–2643, https://doi.org/10.1038/ismej.2017.119.

[81] Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, Nature 447 (2007) 661–678, https://doi.org/10.1038/nature05911.

[82] R.M. Plenge, M. Seielstad, L. Padyukov, A.T. Lee, E.F. Remmers, B. Ding, et al., TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study, N. Engl. J. Med. 357 (2007) 1199–1209, https://doi.org/10.1056/NEJMoa073491.

[83] R.M. Plenge, C. Cotsapas, L. Davies, A.L. Price, P.I.W. de Bakker, J. Maller, et al., Two independent alleles at 6q23 associated with risk of rheumatoid arthritis, Nat. Genet. 39 (2007) 1477–1482, https://doi.org/10.1038/ng.2007.27.

[84] C. Liu, F. Batliwalla, W. Li, A. Lee, R. Roubenoff, E. Beckman, et al., Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis, Mol Med Camb Mass 14 (2008) 575–581, https://doi.org/10.2119/2008-00056 (Liu).

[85] A. Julià, J. Ballina, J.D. Cañete, A. Balsa, J. Tornero-Molina, A. Naranjo, et al., Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility, Arthritis Rheum. 58 (2008) 2275–2286, https://doi.org/10.1002/art.23623.

[86] S. Raychaudhuri, E.F. Remmers, A.T. Lee, R. Hackett, C. Guiducci, N.P. Burtt, et al., Common variants at CD40 and other loci confer risk of rheumatoid arthritis, Nat. Genet. 40 (2008) 1216–1223, https://doi.org/10.1038/ng.233.

[87] J. Cui, K.E. Taylor, A.L. Destefano, L.A. Criswell, E.S. Izmailova, A. Parker, et al., Genome-wide association study of determinants of anti-cyclic citrullinated peptide antibody titer in adults with rheumatoid arthritis, Mol Med Camb Mass 15 (2009) 136–143, https://doi.org/10.2119/molmed.2009.00008.

[88] P.K. Gregersen, C.I. Amos, A.T. Lee, Y. Lu, E.F. Remmers, D.L. Kastner, et al., REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis, Nat. Genet. 41 (2009) 820–823, https://doi.org/10.1038/ng.395.

[89] Y. Kochi, Y. Okada, A. Suzuki, K. Ikari, C. Terao, A. Takahashi, et al., A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility, Nat. Genet. 42 (2010) 515–519, https://doi.org/10.1038/ng.583.

[90] E.A. Stahl, S. Raychaudhuri, E.F. Remmers, G. Xie, S. Eyre, B.P. Thomson, et al., Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci, Nat. Genet. 42 (2010) 508–514, https://doi.org/10.1038/ng.582.

[91] L. Padyukov, M. Seielstad, R.T.H. Ong, B. Ding, J. Rönnelid, M. Seddighzadeh, et al., A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis, Ann. Rheum. Dis. 70 (2011) 259–265, https://doi.org/10.1136/ard.2009.126821.
[92] J. Freudenberg, H.-S. Lee, B.-G. Han, H.D. Shin, Y.M. Kang, Y.-K. Sung, et al., Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci, Arthritis Rheum. 63 (2011) 884–893, https://doi.org/10.1002/art.30235.
[93] C. Terao, R. Yamada, K. Ohmura, M. Takahashi, T. Kawaguchi, Y. Kochi, et al., The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population, Hum. Mol. Genet. 20 (2011) 2680–2685, https://doi.org/10.1093/hmg/ddr161.
[94] H.J. Hu, E.H. Jin, S.H. Yim, S.Y. Yang, S.H. Jung, S.H. Shin, et al., Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis, Exp. Mol. Med. 43 (2011) 613–621, https://doi.org/10.3858/emm.2011.43.11.068.
[95] Y. Okada, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, et al., Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population, Nat. Genet. 44 (2012) 511–516, https://doi.org/10.1038/ng.2231.
[96] J. Wang, A.T. Bansal, M. Martin, S. Germer, R. Benayed, L. Essioux, et al., Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis, Pharmacogenomics J. 13 (2013) 235–241, https://doi.org/10.1038/tpj.2012.8.
[97] S.B. Krintel, G. Palermo, J.S. Johansen, S. Germer, L. Essioux, R. Benayed, et al., Investigation of single nucleotide polymorphisms and biological pathways associated with response to TNFα inhibitors in patients with rheumatoid arthritis, Pharmacogenet Genom. 22 (2012) 577–589, https://doi.org/10.1097/FPC.0b013e3283544043.
[98] K. Myouzen, Y. Kochi, Y. Okada, C. Terao, A. Suzuki, K. Ikari, et al., Functional variants in NFKBIE and RTKN2 involved in activation of the NF-κB pathway are associated with rheumatoid arthritis in Japanese, PLoS Genet. 8 (2012) e1002949, https://doi.org/10.1371/journal.pgen.1002949.
[99] S. Eyre, J. Bowes, D. Diogo, A. Lee, A. Barton, P. Martin, et al., High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis, Nat. Genet. 44 (2012) 1336–1340, https://doi.org/10.1038/ng.2462.
[100] J. Cui, E.A. Stahl, S. Saevarsdottir, C. Miceli, D. Diogo, G. Trynka, et al., Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis, PLoS Genet. 9 (2013) e1003394, https://doi.org/10.1371/journal.pgen.1003394.
[101] S. Negi, G. Juyal, S. Senapati, P. Prasad, A. Gupta, S. Singh, et al., A genome-wide association study reveals ARL15, a novel non-HLA susceptibility gene for rheumatoid arthritis in North Indians, Arthritis Rheum. 65 (2013) 3026–3035, https://doi.org/10.1002/art.38110.
[102] G. Orozco, S. Viatte, J. Bowes, P. Martin, A.G. Wilson, A.W. Morgan, et al., Novel rheumatoid arthritis susceptibility locus at 22q12 identified in an extended UK genome-wide association study, Arthritis Rheumatol Hoboken NJ 66 (2014) 24–30, https://doi.org/10.1002/art.38196.
[103] L. Bossini-Castillo, C. de Kovel, H. Kallberg, R. van 't Slot, A. Italiaander, M. Coenen, et al., A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides, Ann. Rheum. Dis. 74 (2015) e15, https://doi.org/10.1136/annrheumdis-2013-204591.
[104] K. Kim, S.-Y. Bang, H.-S. Lee, S.-K. Cho, C.-B. Choi, Y.-K. Sung, et al., High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci, Ann. Rheum. Dis. 74 (2015) e13, https://doi.org/10.1136/annrheumdis-2013-204749.
[105] S. Senapati, S. Singh, M. Das, A. Kumar, R. Gupta, U. Kumar, et al., Genome-wide analysis of methotrexate pharmacogenomics in rheumatoid arthritis shows multiple novel risk variants and leads for TYMS regulation, Pharmacogenet Genom. 24 (2014) 211–219, https://doi.org/10.1097/FPC.0000000000000036.
[106] L. Jiang, J. Yin, L. Ye, J. Yang, G. Hemani, A.-J. Liu, et al., Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans, Arthritis Rheumatol Hoboken NJ 66 (2014) 1121–1132, https://doi.org/10.1002/art.38353.
[107] N. Govind, A. Choudhury, B. Hodkinson, C. Ickinger, J. Frost, A. Lee, et al., Immunochip identifies novel, and replicates known, genetic risk loci for rheumatoid arthritis in black South Africans, Mol Med Camb Mass 20 (2014) 341–349, https://doi.org/10.2119/molmed.2014.00097.
[108] D.P.C. de Rooy, R. Tsonaka, M.L.E. Andersson, K. Forslind, A. Zhernakova, M. Frank-Bertoncelj, et al., Genetic factors for the severity of ACPA-negative rheumatoid arthritis in 2 cohorts of early disease: a genome-wide study, J. Rheumatol. 42 (2015) 1383–1391, https://doi.org/10.3899/jrheum.140741.
[109] K. Honne, I. Hallgrímsdóttir, C. Wu, R. Sebro, N.P. Jewell, T. Sakurai, et al., A longitudinal genome-wide association study of anti-tumor necrosis factor response among Japanese patients with rheumatoid arthritis, Arthritis Res. Ther. 18 (2016) 12, https://doi.org/10.1186/s13075-016-0920-6.
[110] A. Julià, F. Blanco, B. Fernández-Gutierrez, A. González, J.D. Cañete, J. Maymó, et al., Identification of IRX1 as a risk locus for rheumatoid factor positivity in rheumatoid arthritis in a genome-wide association study, Arthritis Rheumatol Hoboken NJ 68 (2016) 1384–1391, https://doi.org/10.1002/art.39591.
[111] A. Márquez, L. Vidal-Bralo, L. Rodríguez-Rodríguez, M.A. González-Gay, A. Balsa, I. González-Álvaro, et al., A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus, Ann. Rheum. Dis. 76 (2017) 286–294, https://doi.org/10.1136/annrheumdis-2016-209436.
[112] R. Saxena, R.M. Plenge, A.C. Bjonnes, H.S. Dashti, Y. Okada, W. Gad El Haq, et al., A multinational arab genome-wide association study identifies new genetic associations for rheumatoid arthritis, Arthritis Rheumatol Hoboken NJ 69 (2017) 976–985, https://doi.org/10.1002/art.40051.
[113] J. Bluett, S.-A. Owen, J. Massey, A. Alfirevic, M. Pirmohamed, D. Plant, et al., HLA-A 31:01 is not associated with the development of methotrexate pneumonitis in the UK population: results from a genome-wide association study, Ann. Rheum. Dis. 76 (2017) e51, https://doi.org/10.1136/annrheumdis-2017-211512.
[114] Y.B. Joo, Y. Park, K. Kim, S.-Y. Bang, S.-C. Bae, H.-S. Lee, Association of CD8+ T-cells with bone erosion in patients with rheumatoid arthritis, Int J Rheum Dis 21 (2018) 440–446, https://doi.org/10.1111/1756-185X.13090.
[115] W.-H. Wei, S. Viatte, T.R. Merriman, A. Barton, J. Worthington, Genotypic variability based association identifies novel non-additive loci DHCR7 and IRF4 in sero-negative rheumatoid arthritis, Sci. Rep. 7 (2017) 5261, https://doi.org/10.1038/s41598-017-05447-1.
[116] Y.B. Joo, Y. Kim, Y. Park, K. Kim, J.A. Ryu, S. Lee, et al., Biological function integrated prediction of severe radiographic progression in rheumatoid arthritis: a nested case control study, Arthritis Res. Ther. 19 (2017) 244, https://doi.org/10.1186/s13075-017-1414-x.
[117] R. Shinebaum, V.C. Neumann, E.M. Cooke, V. Wright, Comparison of faecal florae in patients with rheumatoid arthritis and controls, Br. J. Rheumatol. 26 (1987) 329–333.
[118] E. Eerola, T. Möttönen, P. Hannonen, R. Luukkainen, I. Kantola, K. Vuori, et al., Intestinal flora in early rheumatoid arthritis, Br. J. Rheumatol. 33 (1994) 1030–1038.
[119] J. Vaahtovuo, E. Munukka, M. Korkeamäki, R. Luukkainen, P. Toivanen, Fecal microbiota in early rheumatoid arthritis, J. Rheumatol. 35 (2008) 1500–1505.
[120] X. Liu, Q. Zou, B. Zeng, Y. Fang, H. Wei, Analysis of fecal Lactobacillus community structure in patients with early rheumatoid arthritis, Curr. Microbiol. 67 (2013) 170–176, https://doi.org/10.1007/s00284-013-0338-1.

2.2 Thesis Aims

The aims to be addressed within this thesis are presented below.

Overarching aim:

1) *Contribute to understanding of the link between the gut commensal microbiota and RA.*

Specific aims:

2) *Apply appropriate methods to the study of the commensal microbiota in RA.*

3) *Investigate the gut microbiota association with RA with regards to host genetic factors.*

4) *Investigate the oral microbiota association with RA with regards to host genetic factors.*

Chapter 3

General Methods 1: Generation of ASVs in TwinsUK

Within this chapter I introduce and describe in detail the core methods used in this thesis. I first introduce the TwinsUK cohort, the data from which has been used extensively throughout the thesis.

I introduce the concept of, and analysis for 16S rRNA gene sequence microbiome data, hereafter referred to as 16S data. I detail the overall steps for microbiome analyses, with particular depth regarding the processing of raw sequencing data: I applied a novel method for the Department of Twin Research for processing of the raw sequencing data to generate the taxon count table which comprises the core 16S microbiome dataset, used for all downstream analyses.

I demonstrate the use of heritability estimation to assess resolution of TwinsUK amplicon sequence variants (ASVs) compared to operational taxonomic units (OTUs), following on from the method applied by Matthew Jackson with regards to OTU generation methods, described in Jackson *et al.* (2016).

3.1 TwinsUK cohort

The TwinsUK cohort is the largest adult twin registry in the UK, with 14,838 participants, recruited during the past 27 years (Verdi et al. 2019). The TwinsUK cohort, is appropriate for the study of rheumatoid arthritis: the majority of participants are female (82%) and the

mean age of participants is 59. The cohort is deeply phenotyped, using data collected during ongoing visits to the Clinical Research Facility at St Thomas' hospital, managed by King's College London. An array of associated 'omics' datasets have been generated (Verdi et al. 2019) Volunteers within the TwinsUK are particularly engaged and forthcoming in participating in ongoing research.

3.2 Background: Facilitation of microbiome studies by development of next generation sequencing platforms

Until recently, microbial studies were limited to culture dependent techniques. Substantial technological advancements have facilitated progression in the characterisation of micro-organisms from culturing, to genomic sequencing. Following the success of the human genome project, massively parallel sequencing systems were developed to improve sequencing efficiency (Chial 2008). Scientific advances in genomic sequencing proceeded through three major technical revolutions, comprising first generation sequencing - whole genome shotgun sequencing, next generation high throughput sequencing (NGS) and third generation of sequencing, single molecule long read sequencing (Cao et al. 2017). Since the natural economic progression leading to reduction in cost following the initial development of NGS circa 2007, there has been exponential interest in the application of these techniques to study the microbiome. Subsequently the microbiome field of study has rapidly expanded, constituting a tipping point concerning understanding of human physiological function in health and disease. This is described in further detail in Chapter 1: Introduction. The main platforms which have been developed for high throughput sequencing are, in order of evolution, ROCHE 454, SOLiD, Ion Torrent, Illumina Hi-Seq and Illumina MiSeq (Pillai et al. 2017).

3.2.1 16S rRNA gene sequencing

The most widely employed method for measuring the microbiome composition of biological samples is the sequencing of the 16S ribosomal RNA (rRNA) marker gene. The efficacy of this approach and lower cost relative to metagenomics methods, underlies its popularity. The 16S rRNA gene encodes for a component of the 30s small subunit of the prokaryotic ribosome (Figure 3.1) and is therefore ubiquitous amongst bacteria: it is a pre-requisite for synthesis of all bacterial proteins. Hence it is an ideal marker gene candidate. The structure of the 16S rRNA gene (described below) confers suitability for use as a taxonomic marker.

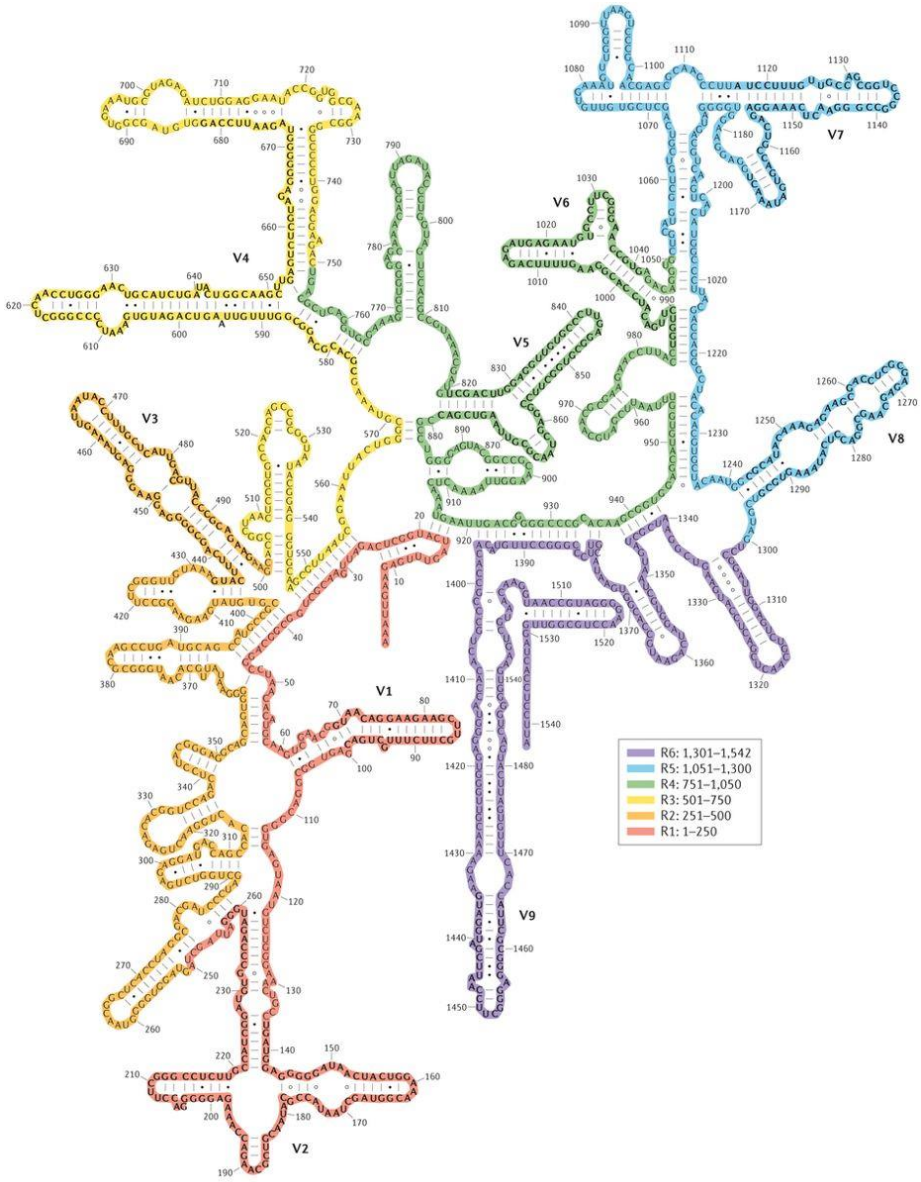


Figure 3.1 Secondary structure of 16S rRNA of Escherechia coli
Adapted from Yarza et al. (2014)

Within the 16S rRNA, there are 9 hypervariable regions, the sequence of which is (predominantly) taxon specific, and each is flanked by highly conserved, non taxon-specific regions. Following DNA extraction from samples, the use of conserved region-specific primers allows amplification with polymerase chain reaction (PCR), of the targeted variable regions (Figure 3.2).

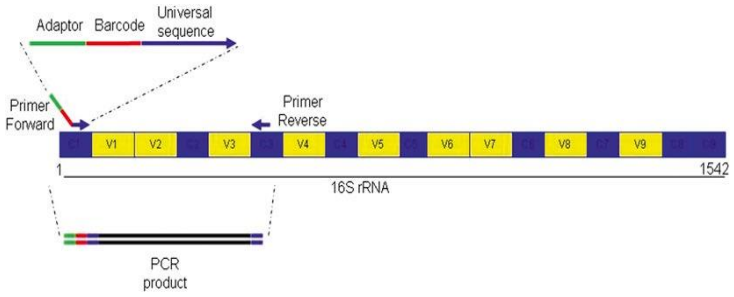


Figure 3.2 16S rRNA gene sequence showing conserved and hypervariable regions
Adapted from Del Chierico, *et al.* (2015)

PCR generated amplicons are sequenced using a next generation sequencing platform. Initially Roche 454 platforms were most widely used, but have now been superseded by Illumina MiSeq. Of the 9 hypervariable regions V1 – V9, V4 is the most frequently used within 16S studies, and is used in TwinsUK. The V4 region is 253 base pairs in length, and full-length sequencing can be achieved using paired (forward and reverse) reads generated on an Illumina MiSeq. Once sequence reads are generated, pre-processing steps are undertaken. At this stage, a count table is generated (generation of OTUs or ASVs, described below), and sequence groups are compared to a reference ribosomal RNA gene database such as Greengenes, RDP or SILVA, in order to identify taxa present within a sample.

3.2.2 Generation of sequence counts: ASVs

Sequencing data must be converted to a table of taxonomic counts present per sample, prior to statistical analysis. Traditionally, this has been achieved via generation of OTUs, in which reads are grouped together according to a similarity threshold (Edgar 2017). More recently, higher resolution methods have been developed: ASVs (Callahan et al. 2016, 2017). The principle of ASVs is that sequencing errors are resolved to reveal the original biological sequence, allowing exact matching of sequences as opposed to grouping

according to a similarity threshold. This overcomes some limitations of OTU grouping based methods (Callahan et al. 2017). Sequences which would otherwise have been contained within the same OTU - but are from different taxa - can be separated and more accurately identified. The broad concept of this is illustrated in Figure 3.3. Sequences which vary by just one nucleotide may also be (arbitrarily) assigned to different OTUs; this is not the case with ASVs. Additionally, as the original biological sequence has been resolved, it is possible to assign taxonomy at species level in a proportion of the data. A final advantage of ASVs representing a true biological entity is that they can be directly compared between datasets (Figure 3.4).

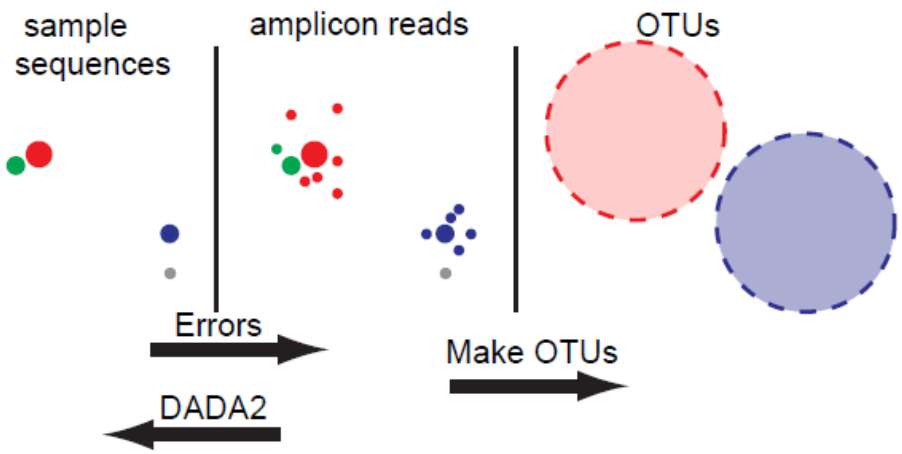


Figure 3.3 Graphical representation of the theory of increased resolution of ASVs versus OTUs
Adapted from Callahan *et al.* (2016). In the centre are shown amplicon reads with different sequences. Each colour of amplicon reads in the centre partition are derived from the same bacterial species. The blue and the grey, and the red and the green are sequentially similar to each other (>97% similarity) but represent different taxa. On the right, the similar sequences are grouped together to form two OTUs, derived from the amplicon reads from four different taxa. On the left, four ASVs have been devised representing the four taxa from which the amplicon reads originated.

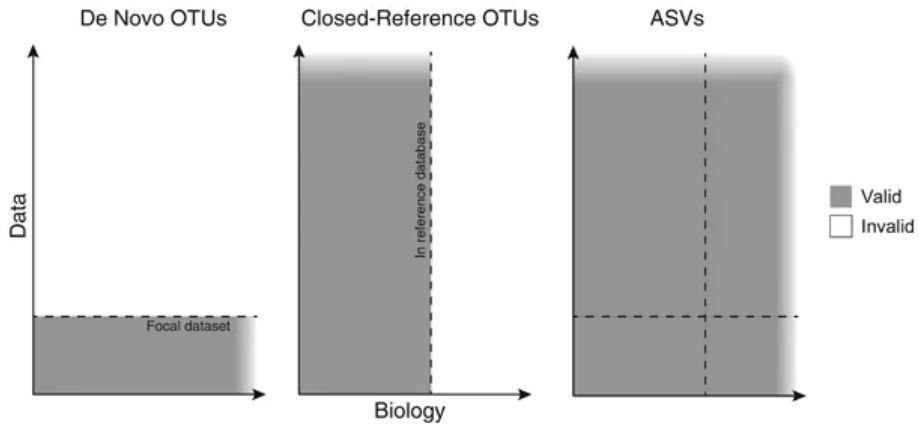


Figure 3.4 Validity of ASVs and OTUs within and between datasets
Adapted from Callahan *et al.* (2016). On the right, ASVs represent a true biological entity, and are therefore directly comparable between different datasets. Closed reference OTUs are limited in the biological information captured owing to initial direct comparison with incomplete reference databases during generation. De-novo OTUs have higher bandwidth for capturing biological information compared to closed reference OTUs but are not directly comparable between datasets.

ASVs are generated by using the sequencing error pattern of reads generated by next generation sequencing to infer the original biological sequence. The method, developed by Callahan *et al.* (2016), is optimised for Illumina MiSeq V4 16S rRNA data, though can be applied to other data types such as fungal nuclear ribosomal internal transcribed spacer (ITS).

Due to variation in errors between sequencing runs, each run is dealt with separately. Reads (organised into per sample forward and reverse files) are first trimmed, as quality deteriorates toward the ends of reads secondary to reagent degradation. The data is then de-noised using the following 5-step process.

First, sequences are first dereplicated (matched with identical sequences and assigned an abundance value) to increase computational efficiency. Second, sequences are compared, and at this stage the data-specific error pattern is modelled. Third, sequences are partitioned by applying the divisive partitioning algorithm: the most abundant unique sequence is used as the central comparator, and data is divided until all unique sequences

within the partition are statistically deemed to have likely originated from the central unique sequence, statistical likelihood is determined via calculation of the sequence abundance P value (Figure 3.5). The sequence abundance P value also identifies (leading to exclusion of) reads which are likely to be sequencing errors. By default, all singleton reads will be identified as potential sequencing errors, as it is not possible to determine the likelihood that a singleton read represents a biological entity. The fourth step is error model parameterisation, using non-parametric polynomial based Loess modelling and application to resolve sequencing errors to finally reveal (theoretically) the original biological sequences of the reads. Fifth, forward and reverse reads are joined (and because errors have been resolved, exact base pair matches are made), and ASVs from different sequencing runs are then merged.

After denoising, chimera removal is undertaken. Chimeras in this context are artefacts of PCR (discussed in detail in Chapter 8). A conservative chimera removal method is applied. Finally, taxonomy of the ASVs is assigned, using a reference database such as SILVA or Greengenes (Table 3.1).

3.2.3 Evidence of comparative efficacy of ASVs: Benchmarking

Callahan and colleagues undertook comprehensive benchmarking of ASVs compared to other methods for inference of sample composition using 16S rRNA gene sequencing (Callahan et al. 2016). Using mock reference bacterial communities for which the taxonomic composition was known, they compared the performance of ASVs generated using DADA2 to other algorithms which produce OTUs: UPARSE, MED, Mothur and QIIME. The same variable region as is used within TwinsUK (V4) was sequenced using also the same primers (F515 and 806R; Section 3.4.1). Also in accordance with TwinsUK, 250 base pair pared end reads were used. Three types of mock community were used – balanced, HMP and extreme. In all three datasets, DADA2 demonstrated higher sensitivity

and specificity compared to other methods. The HMP dataset is the most comparable to the data in this thesis, consisting of intestinal bacteria. When comparing efficacy in identification of sequences for all types of mock community across each algorithm, DADA2 consistently outperformed other methods (Figure 3.5).

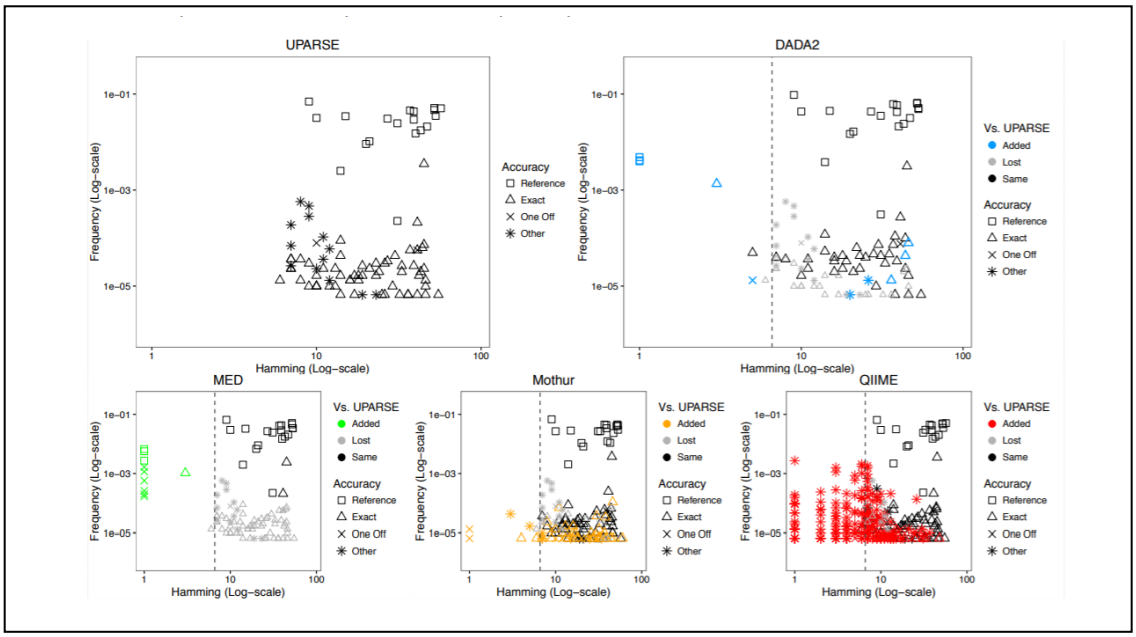


Figure 3.5 Comparison of output sequences inferred from HMP mock community by HMP, UPARSE, DADA2, MED, Mothur (average-linkage) and QIIME (Uclust)
On the Y axis is plotted frequency of output sequences from the HMP merged dataset. On the Y axis is Hamming distance from each sequence to its nearest more-abundant neighbour sequence. UPARSE is used as a baseline with which to compare the methods. Algorithms largely concur (black) in identifying sequences that are abundant and very different from other sample sequences. However - DADA2 detects additional variation (blue), which is accurate (square or circle). MED does also detect some fine-scale variation (green), but at the cost of some additional false positives, and MED demonstrates more lost data as it does not detect low abundance sequences (grey). Mothur (orange) and QIIME (red) also both report a substantial number of additional spurious sequences (which was not the case for DADA2).
Adapted from Callahan et al. (2016).

3.3 TwinsUK 16S data

3.3.1 Sample collection and pre-sequencing steps

Faecal samples for participants within the TwinsUK cohort were collected during clinical visits from 2010-2015. Samples were produced during the two days prior to the visit in 15mm canonical tubes and stored in a refrigerator for up to two days prior to clinical visit. When bowel movement was not possible, a faecal sample was produced at a later date as soon as possible and posted to the department of twin research in a sealed ice pack. Upon arrival, samples were frozen at -80 degrees C.

Frozen samples were shipped on dry ice to Cornell University, where DNA extraction, polymerase chain reaction (PCR) and sequencing was undertaken. Prior to DNA extraction, samples were removed from the freezer. Physical abrasion of samples with bead beating was included in the protocol in order to improve yield from gram positive bacteria, which have a robust cell wall. Genomic DNA extraction was performed on 100 mg sample aliquots using the PowerSoil® - htp DNA isolation kit (MoBio Laboratories Ltd, Carlsbad, CA). PCR amplification of the 16S rRNA gene was applied in triplicate. Primers targeting the V4 hypervariable region were used: F515 and 806R, as described in Caporaso et al. (Caporaso et al. 2011). PCR replicates were combined and purified using a magnetic bead system (Mag-Bind® EZPure, Omega Bio-Tek, Norcross, GA). Quantification of PCR amplicons was undertaken using the QuantiT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA), followed by combining aliquots of amplicons (at equal masses) at a final concentration of approximately 15 ng/μl. DNA was sequenced using the Illumina MiSeq platform, generating 250 base pair (bp) paired-end reads, at Cornell Biotechnology Resource Centre Genomics Facility.

3.3.2 Generation of ASVs in TwinsUK

Generation of ASVs is a multiple step process, described above in **Section 3.2.2**. In the present section I provide a description of the preparatory computational steps and detail

the problem-solving steps specific to the application of the technique to TwinsUK 16S data (Table 3.1).

The first step was to produce demultiplexed FASTQ files containing read sequence, sample ID and read quality information. Reads for each sequencing run (n=35) were first demultiplexed (assigned to samples according to identifier barcodes; Figure 3.6) in QIIME with minimum thresholding parameters as specified in the DADA2 protocol (Figure 3.7).

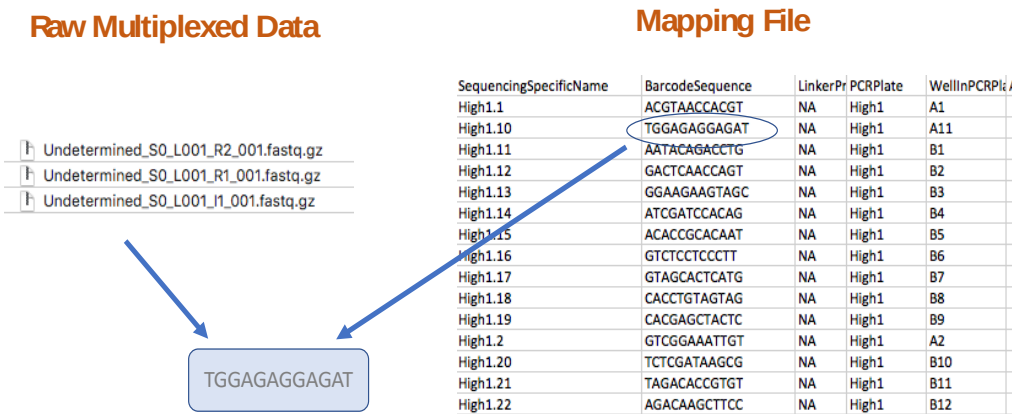


Figure 3.6 Representation of barcodes linking multiplexed raw sequencing and meta-data
12bp Golay barcode linking raw sequencing data with the metadata for each sample, contained within the mapping file.

After demultiplexing, the forward and reverse file for the sequencing run was split according to sample ID, producing separate forward and reverse files for each sample within the sequencing run in QIIME using the command split_libraries_fastq.py (Figure 3.7).

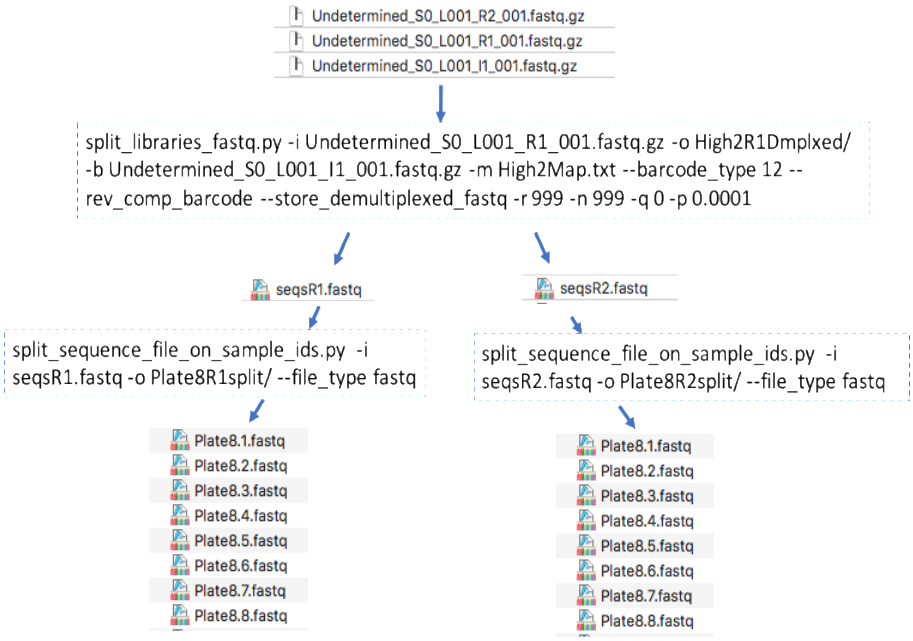


Figure 3.7 Generation of sequencing files for input to DADA2 package

In total there were 35 sequencing runs requiring independent custom filtering thresholds, due to variation in quality of the reads between runs. Following filtering, reads should ideally be of good quality throughout (Figure 3.8). The thresholds used for each sequencing run are given in the ASV script in Appendix A.

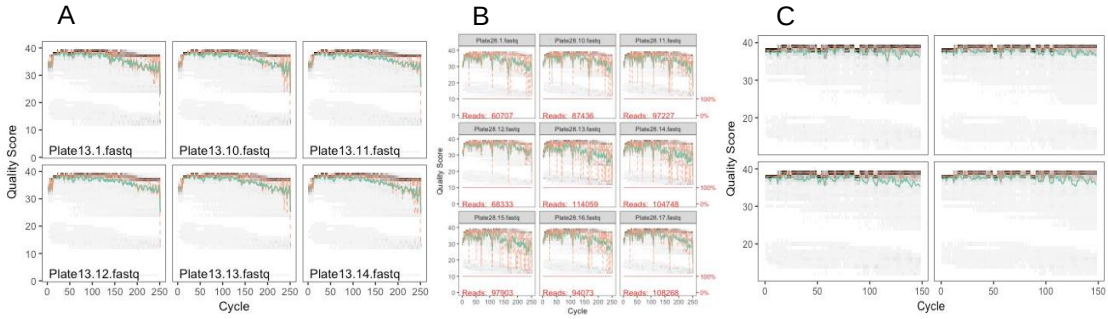


Figure 3.8 Quality plots of sequencing reads prior to and post trimming
On the y axis is shown the phred quality score of the sequencing read. On the x axis is shown the number of bases. In A, higher quality reads from an example sequencing run are shown. In B, an example of poorer quality sequencing reads are shown, containing fluctuation in quality throughout the read, as opposed to towards the end of the read, as would be usual. In C, an example of trimmed reverse read is shown, with the spurious quality portion removed, which was at the end of the read.

ASVs were generated as described above in Section 3.2.2, and taxonomy was assigned using the SILVA 1.3.2 database. A phylogenetic tree was generated for the ASVs using the Phangorn package in the programming language R with default parameters (Schliep 2011). During the generation of the ASVs there were several issues which required resolution. In particular, there were problems with the barcodes and overall quality of the reads (Figure 3.8). These problem-solving steps were the rate limiting factor for this work. The troubleshooting process for each issue is detailed below in Table 3.1.

Issue	Description /Problem Solving	Solution
Demultiplexing	Previous work on this data began from demultiplexed FASTA files. For DADA2, FASTQ files containing quality score information is required. The correct demultpiplexing method and commands for the data needed to be devised.	Troubleshooting through trial and error was required. An additional flag: --Rev_comp_barcode was added to the script (identified by Maria Stalteri).
Plate 16 barcodes	Demultiplexing was not successful for plate 16. The barcodes in the mapping file were incorrect for this sequencing run.	I wrote an R script to extract the barcodes from existing FASTA files for the same data, and insert the correct barcodes into the mapping file (Appendix A). Demultiplexing using the updated barcodes was successful
Only reverse read available – forward read missing	For one sequencing run, only the reverse sequencing file was available.	In order to avoid the need to remove this sequencing run entirely, I amended the script to generate ASVs from the reverse read only.

Trimming parameters	Sequencing information deteriorates in quality at the ends of the read, as the reagents degrade. It is therefore suggested in the manual to trim the reads. Little guidance was available with regards to choosing the parameters.	There is a balance between removing spurious quality data, and maintaining sufficient read length and overlap between forward and reverse reads.
Sequence runs crashing: quality of reads too low for DADA2 algorithm	Some samples were of particularly low quality, such as is depicted in example B in Figure 3.8.	For affected sequencing runs, I sequentially removed the worst quality samples until the script was able to run.
Only reverse read available – forward read missing	For one sequencing run, only the reverse read was available.	I adapted the ASV script to generate ASVs from only the reverse read for this sequencing run.
Confirming that spurious sequences were successfully identified and removed.	Within the DADA2 workflow the chimera removal step serves to identify not just chimeras (non-biological artefacts of PCR), but also whether there were non-biological instertions (e.g. adapters) present in the data. If present, this would indicate that the data required further steps of adapter removal, before re-running the workflow.	I checked the proportion of reads removed at the chimera removal step using the following calculation: sum(seqtab.nochim)/sum(seqtab) This identified that 10% of reads were removed. I contacted the creator of DADA2 (Ben Callahan) regarding this, and he confirmed that this is comfortably within the expected parameters.
Choice of referencing database	Taxonomic classifications of 16S rRNA gene sequences is usually undertaken using one of four	The SILVA database has been most recently updated compared to the other resources. In the SILVA database, where species assignment is not possible due to

	resources: SILVA, RDP, Greengenes or NCBI.	constraints of the data, further efforts have been made to sub-classify sequences in a level between genus and species, by assigning according to group of species, increasing resolution of taxon assignment. The SILVA database was therefore selected as the most appropriate database.
Missing taxonomies	A greater than expected number of sequence taxonomies were unassigned initially.	I identified that in some instances ASVs are the reverse complement of the reference, and I re-ran the step adding an additional flag: tryRC == TRUE.
Some ASVs not merging.	Some ASVs not merging when sequencing runs were combined.	I identified that this was due to the differing sequence lengths, and I re-ran the step adding an additional flag: collapseNoMismatch.

Table 3.1 Problem solving undertaken during generation of ASVs in TwinsUK 16S data

3.4 Efficacy of ASVs

Heritability estimation provides the scope to compare resolution within and between OTU and ASV based methods. The hypothesis is that methods which perform better are more likely to produce units (ASVs or OTUs) which represent a true biological entity, and that this confers higher detectable heritability.

3.4.1 Calculation of narrow sense heritability estimates for ASVs

Heritability estimation calculations utilise correlation of a phenotype between monozygotic (MZ) and dizygotic (DZ) twin pairs, with respect to the average expected shared genetic and environmental within and between pair variance. Narrow sense heritability is the ratio of additive genetic variance to the total phenotypic variance. Dizygotic twins share on average 50% of their heritable genetic material, whilst monozygotic twins share 100%. Classical twin modelling relates the observed variance of a trait to the effects of unobserved (latent) factors, which are linked to functional effects of genes (A), shared environment (C) and unique individual-specific environment (E) (Figure 3.9). A key (but sometimes criticised) assumption of the model is that DZ and MZ twin pairs share the same level of environmental influence. Overall, it is an elegant technique for the purpose of investigating genetic, shared environmental and unique environmental influences.

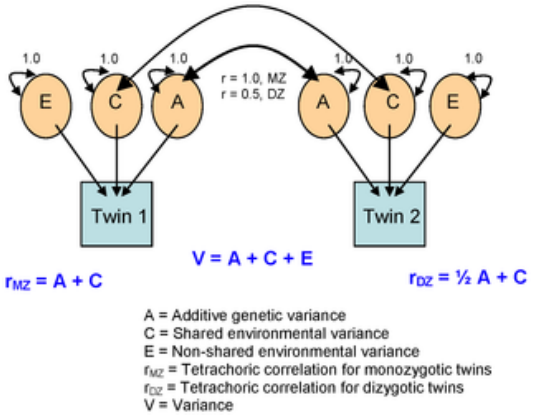


Figure 3.9 ACE model for calculating genetic and environmental influence from MZ and DZ twin correlations
Adapted from (Bunyavanich et al. 2013)

Prior to model fitting, the powerTransform package in R was used to estimate a Box–Cox transform lambda producing approximately normally distributed residuals from a linear model with OTU abundance as a response variable and gender, age, sequencing run, sequencing depth, postal or non-postal sample collection, technician responsible for DNA extraction and technician responsible for preparing the sequencing plate as predictor

variables. This was undertaken for each ASV and the transformed residuals were used as input for heritability estimation.

Maximum likelihood model fitting, adjusting for age, sex and technical factors was applied to estimate A (additive genetic), C (shared environmental), and E (unique environmental) contributions to ASV abundance, from data for monozygotic and dizygotic twins, the using the openMX R package.

Heritability estimates were calculated as described above, at both the ASV level and the genus level. Genus level refers to grouping together ASVs with the same genus level annotation. The genus group annotations from the SILVA database includes subdivisions of ASVs into predicted groups of species, and these were grouped together. These annotations are useful for differentiating ASVs from the same genus, where annotation at species level was not possible due to insufficient sequence resolution or length. There are both advantages and disadvantages to calculation at ASV or genus level. At the ASV level, heritability estimates of multiple ASVs with the same annotation represent independent findings in the context of other heritability estimates within the analysis. However, it is impossible to ascertain whether these are members of the same taxa, or whether there is divergence at the species or strain level. Due to the lower numbers of each ASV compared to the genus group to which it belongs, confidence intervals of heritability estimates at the ASV level will be wider. Heritability estimates at the genus level demonstrate narrower confidence intervals, due to higher power where these grouped ASVs represent the same taxa. This is illustrated in the study by Goodrich *et al.* (2016) in which heritability of OTUs in TwinsUK was calculated in a smaller and subsequently larger sample of the cohort, and in the latter, confidence intervals were narrower (**Figure 3.2.4**). It is also evident in the heritability of ASVs (**Figure 3.2.1** and **Figure 3.2.2**) compared to ASVs grouped at the genus level (**Figure 3.5**).

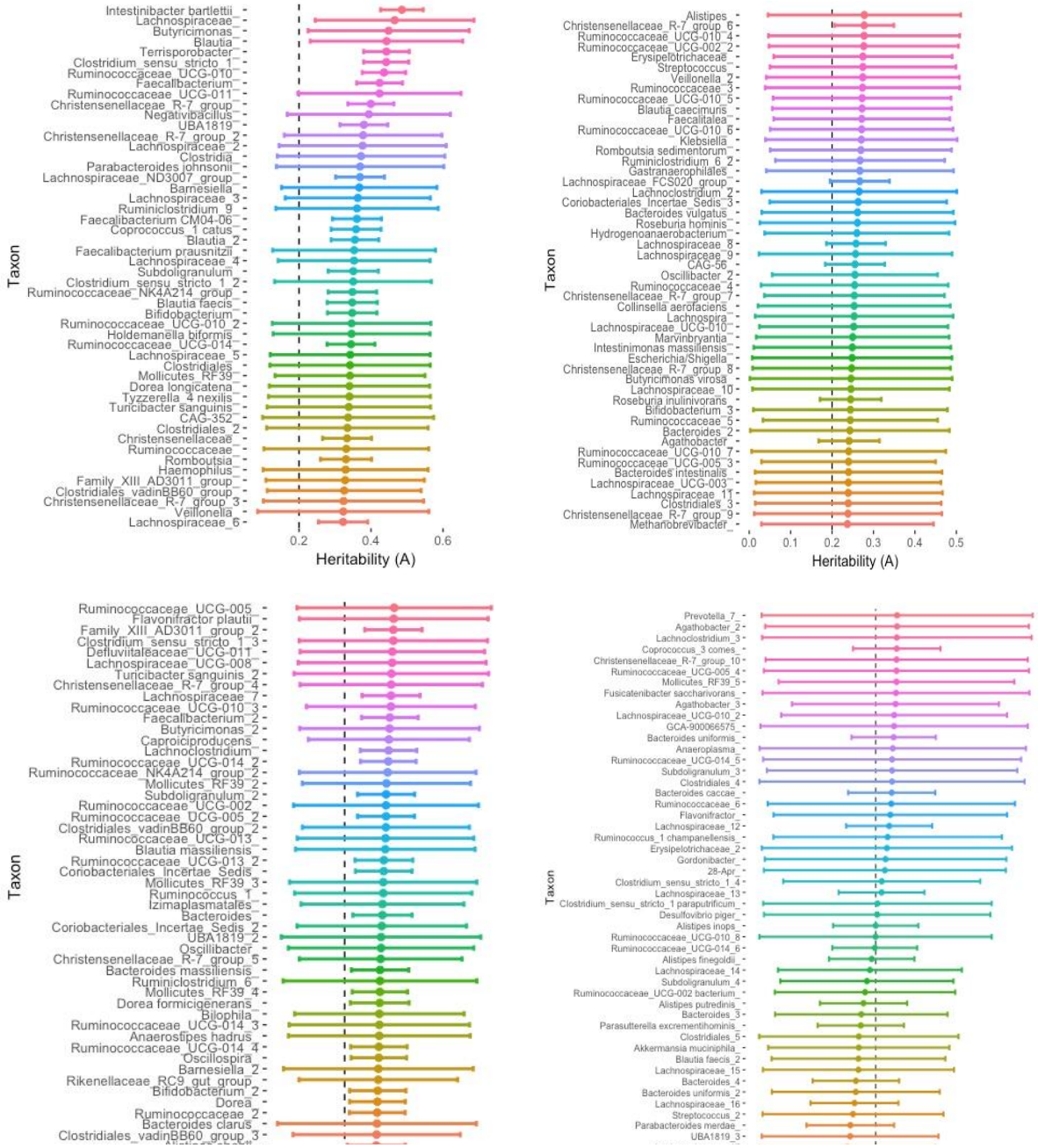


Figure 3.2.1 Heritability of ASVs present in more than 5% of samples.

There were 200 significantly heritable ASVs present in more than 5% of samples. The most heritable ASV was *Intestinibacter bartletti* (49%, 95%CI 43-56).

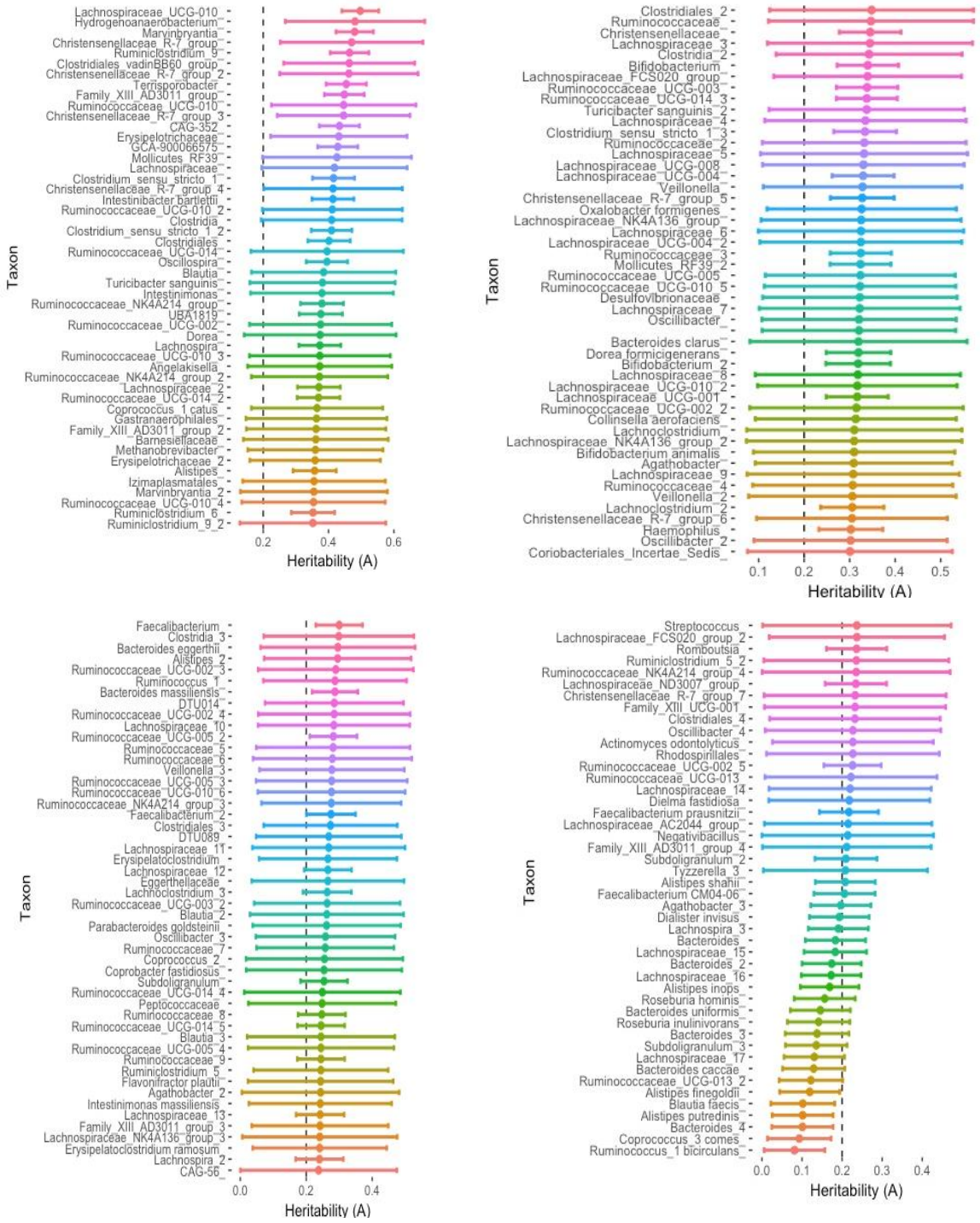


Figure 3.2.2 Significantly heritability ASVs present in more than 15% of samples

There were 159 significantly heritable ASVs present in more than 15% samples. The most heritable ASV was Lachnospiraceae_UCG-10 (49%, 95%CI 44-55)

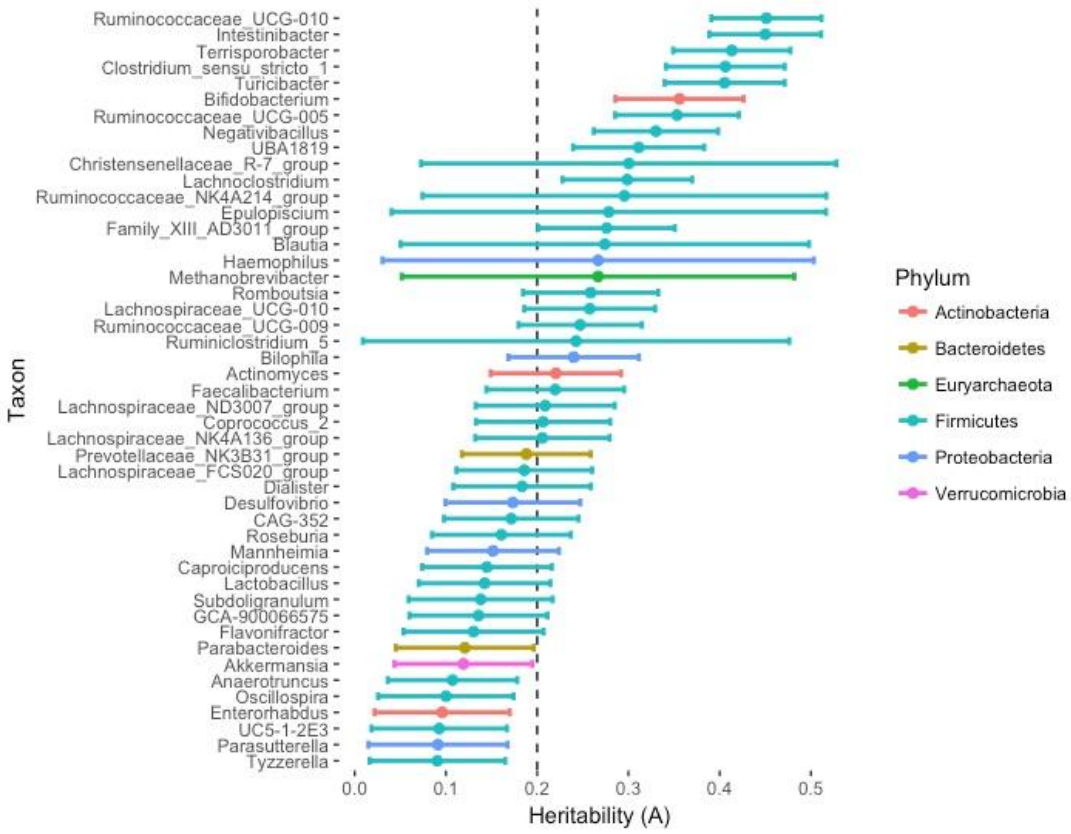


Figure 3.2.3 Significantly Heritable ASVs at genus level, present in more than 5% of samples
At genus level, of ASVs present in more than 5% of samples, 47 were significantly heritable. The most heritable taxon was Ruminococcaceae_UCG-10 (45%, 95%CI 39-51).

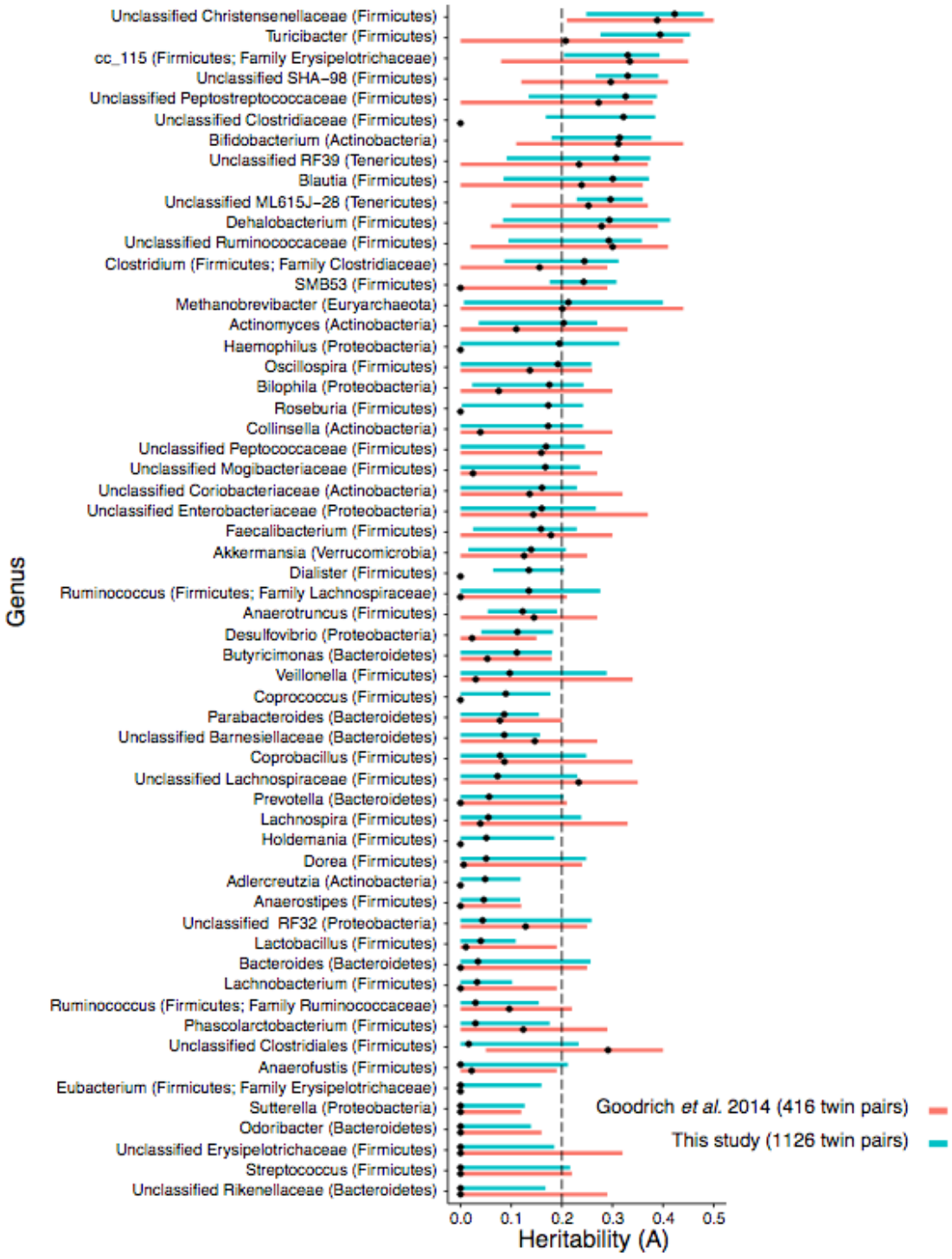


Figure 3.2.4 Heritability of 58 OTUs collapsed to genus level
Adapted from Goodrich *et al.* (2016)

3.4.2 Comparison of narrow sense heritability estimates for ASVs and OTUs

Direct comparison between heritability estimates of OTUs and ASVs is challenging, because of the differences in the methods, and particularly the resolution and grouping of the sequences. In addition, in the ASVs there is a merge step in which ASVs with the same sequence are merged across sequencing runs to include all samples. On the other hand, OTUs are created in the entire dataset, producing a greater number of units which may represent the same taxon compared to ASVs. OTUs also contain taxa within the same OTU which are sequentially 97% similar to each other, the 3% leeway in sequence difference sometimes confers that they represent different species. Apparent heritability of some of the taxa represented as OTUs may rather reflect the increased power afforded by grouping together of sequentially similar taxa. OTUs were considered at the genus level, so although there were more OTUs than ASVs all with the same genus annotation were considered as one entity; The prior comment on power would not apply if individual OTUs had been analysed. Finally, the OTUs have been annotated using the Greengenes database, whereas the ASVs have been annotated using the SILVA database.

Overall, there are four key factors to consider: heritability (A), confidence intervals, sequence clustering differences and success regarding taxonomy assignment.

The ASVs showed higher heritability compared to the OTUs. The most heritable ASV (Rumicoccaceae_UCG_010) was estimated to be 45% heritable whereas the most heritable OTU (Christensellenaceae, genus unclassified) was estimated to be 38% heritable. Rumicoccaceae_UCG_010 was not identified amongst the significantly heritable OTUs, however a genus unclassified OTU from the Ruminococcaceae family was one of the most heritable OTUs (35%) and is potentially the same taxon.

In the OTUs, in the study by Goodrich *et al.,* the most heritable taxa at the genus level was Christensellenaceae, genus unassigned. This was not replicated in the ASV heritability analysis at the genus level, however genus grouped ASVs assigned to Christensellenaceae R7 group were the 10th most heritable taxon. In addition, at the ASV level, many ASVs belonging to the Christensellenaceae family were significantly heritable, it is likely that in the OTUs, some of these are grouped within the same OTU. The equivalent result in ASVs demonstrated that whilst Christensellenaceae is substantially heritable, it is not the most heritable taxon. Further, using OTUs, it was not possible to taxonomically identify the Christensellenaceae OTU beyond family – genus was unassigned. In addition, using ASVs, I have demonstrated that multiple ASVs assigned to the Christensenellaceae family are significantly heritable. As suggested above, it is possible that some of these were grouped into the same OTU, rendering the higher heritability estimate demonstrated for Christensellenaceae in the OTUs.

In the OTUs, the second most heritable taxon was *Turicibacter* with a heritability estimate of 38%. In the ASVs, *Turicibacter* was the 5th most heritable taxon, but this genus showed higher heritability in the ASVs (41%) than in the OTUs. The third and fourth most heritable taxa in the OTUs – Erysipelotrichaceae _Cc115 and SHA-98 were not replicated in the ASVs. As OTUs do not identify novel information compared to ASVs, it is instead likely that some samples containing higher abundance of these taxa were not of high enough quality to successfully produce ASVs and were excluded from the data (Table 3.1). Our collaborator at the University of Stanford, Daniel Sprockett, undertook similar analysis of heritability of his independently created ASVs and replicated this (results not shown).

Amongst the results for the OTU heritability estimates there are 11 OTUs which were unclassified at the genus level. In the ASVs, ASVs which were unclassified at the genus level have been removed, however some OTUs which could not be classified at the genus level were possible to identify using ASVs. This is potentially because the denoising algorithm extrapolates to the original biological sequence before sequencing errors were introduced, and therefore comparison of sequences to the reference database is more successful.

However it could also potentially represent ASVs and OTUs from the same family but different genus.

In the OTUs, Prevotella genus is identified as one of the significantly heritable taxa, whereas in the ASVs, use of the SILVA database has allowed this taxon to be more accurately identified as comprising of Prevotellaceae NK3B31 group and *Prevotella_7.* The latter only just reached significance - the 95% CI was only just greater than 0.

3.4.3 Summary

In summary, ASVs generated in TwinsUK contained novel information to OTUs generated from the same raw sequencing data, supporting the demonstration by Callahan *et al.* (2017) that ASVs offer substanbitally higher resolution. The TwinsUK ASVs contain novel information compared to OTUs, which will be valuable when considering association with phenotypes. Annotation of the ASVs with the SILVA database allowed differentiation of some taxa above genus level in instances where species annotation was not possible. For example, it can be deduced that the heritable *Prevotella* OTU is not *Prevotella_copri;* the heritable *Prevotella* ASV was assigned to *Prevotella_7* . A limitation identified was that ASVs require higher quality sequence reads as input compared to OTUs - some samples were removed as it was not possible to apply the method to them - and this was not the case with OTUs. However, the ASV methods are higher resolution and more accurate. In addition, creating the ASVs afforded development of a greater personal depth of understanding of the method, compared to if I had used data curated by collaborators.

3.4.4 Conclusion

ASVs are the optimal method presently available for the pre-processing of 16S data. The large ASV dataset that I generated will improve resolution of results in the following thesis, in comparison to use of OTUs. Publications in which the ASVs have been utilised are listed in the beginning of the thesis: all non-review articles include analyses of these data. In the following chapter, I consider appropriate statistical methods with regards to analysis of ASVs.

Chapter 4

General Methods 2: Statistical analysis of the microbiota

In this chapter I describe the challenges for statistical analysis of the microbiota as represented by ASVs, and I provide the background for the choice of techniques used in this thesis.

4.1 Core downstream statistical analysis workflow

The ASV counts which represent the core microbiome dataset present a challenge with regards to statistical analysis; There are several schools of thought with regards to the most appropriate approach (Gloor et al. 2017, Weiss et al. 2017). Analysis of microbiome data is a rapidly expanding field. Microbiome analysis holds specific challenges and specialised techniques have been developed to account for the high dimensionality of the data. There are an array of approaches available for each category of analysis. A particular challenge is the selection of an appropriate normalisation technique (discussed below in Section 4.2.2).

The broad measures derived to understand microbiome composition are alpha diversity (within sample) and beta diversity (between sample), and taxonomic associations (Figure 4.1). A schematic of the workflow for understanding association of the microbiota with the phenotypes examined in this thesis is presented in Figure 4.2.

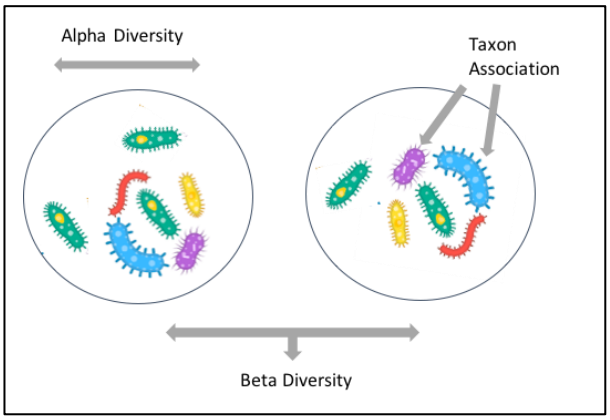


Figure 4.1 Key measures of interest in microbiome studies: Alpha and beta diversity, and taxon associations

Alpha diversity represents in the within-sample microbial diversity, which can be used to compare from one sample to another, whilst beta diversity quantifies the overall dissimilarity between samples. These measures are discussed below in Section 4.2 and Section 4.4, respectively. Taxon associations provide more fine-scale information regarding association of the microbiome with a phenotype and may be present in the absence of detectable diversity differences. In all analyses, variation in sequencing depth between samples must be accounted for (methods described in Section 4.3).

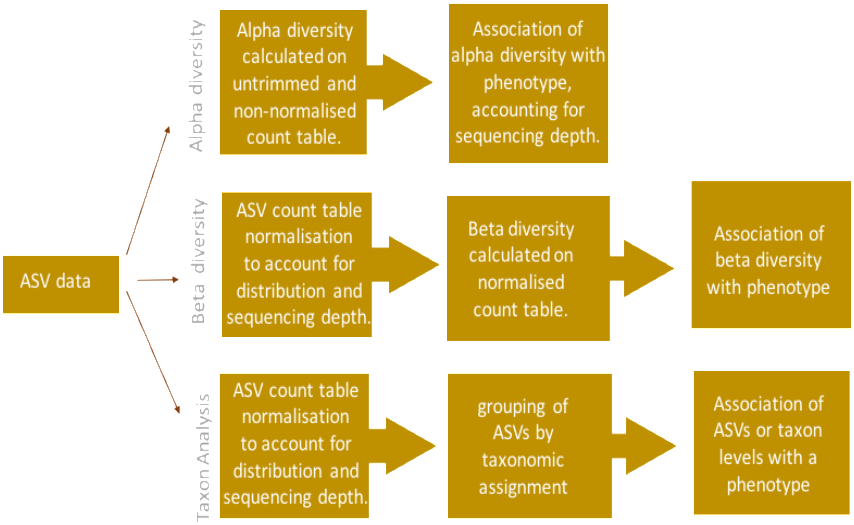


Figure 4.2 Overview of steps undertaken to measure microbiota association with a phenotype.

4.2 Alpha diversity

Alpha diversity metrics capture the microbial diversity within a sample, which can be used to investigate association of sample diversity with a phenotype. Alpha diversity metrics were originally developed by ecologists, but are now widely applied to microbiome studies, as the microbiome represents an ecological community which have been suggested to share many of the principles of classical ecology (Fierer et al. 2012). However this assumption has been criticized (Reese and Dunn 2018). Alteration in alpha diversity of the gut microbiota, and typically a lower alpha diversity, is a frequently demonstrated feature of disease, but is not ubiquitous. Correspondingly, disease treatment with prescription medications is associated with lower gut microbiota alpha diversity (Jackson et al. 2018a). Different alpha diversity metrics demonstrate comparative variation in one or more different aspects of microbial diversity: species richness, species evenness and phylogenetic variation. Species richness refers to the number of different taxa present in the sample, without regard to their frequencies. Evenness refers to the variation or lack thereof in abundance of taxa within a sample (Wagner et al. 2018). Alpha diversity measures include Shannon's diversity index, Chao1, Observed ASVs, Simpson index and Faith's phylogenetic diversity (Shannon and Weaver 1949, Simpson 1949, Chao 1984, Faith 1992).

Shannon diversity (Shannon and Weaver 1949) accounts for both richness and evenness of species. The Shannon index is an information statistics index, and carries the assumptions that all taxa in the ecosystem from which the sample originates are present in the sample, and that they have been randomly sampled. These assumptions are partially violated by high throughput sequencing techniques which rely on primers, this issue is described in greater detail in Chapter 8, Discussion. To calculate the Shannon index, the proportion of each taxa relative to the total number is multiplied by the natural logarithm of the proportion. The product is then summed across species and multiplied by -1.

The Simpson index (Simpson 1949) also takes into account both richness and evenness of species, however it is a dominance index – it gives more weight to common taxa, which are dominant within the sample. This index measures the probability that two organisms from the same sample will be members of the same taxa. Using the Simpson index, presence of rarer taxa with few representatives within the sample would not affect the measure of diversity for that sample – more weight is given to the more abundant taxa. This may be of use for, or a hindrance to the study, depending on the biological question at hand.

Observed ASVs is simply the total number of ASVs present in a sample, and as such is a measure of species richness. Faith's phylogenetic diversity (Faith 1992) is a measure of how phylogenetically diverse taxa are within a sample, and this measure does not incorporate taxon richness or evenness; here, the number branches of the phylogenetic tree linking taxa within the sample are summed. More phylogenetically diverse samples will contain a greater number of linking branches in the phylogenetic tree and will demonstrate higher Faith's phylogenetic diversity. Chao1 (Chao 1984) relies on singleton ASVs or OTUs to calculate the measure. As singletons are not a feature of ASV data owing to the first step for each ASV which is comparison of the two most similar sequences which leads to the primary discarding of singletons, Chao1 has not been applied in this thesis.

In this thesis, Alpha diversity is calculated using untrimmed ASVs (including all ASVs in the sample), and non-normalised count tables, as recommended by McMurdie and Holmes (McMurdie and Holmes 2014). Sequencing depth is accounted for as a fixed-effect covariate when measuring association of alpha diversity with a phenotype using a linear mixed model.

4.3 Normalisation of sequence counts

Prior to application of other techniques to investigate beta diversity or taxonomic associations, the sequence counts must first undergo normalisation. There are two key reasons for this; the first is to transform the distribution of the sequence counts which are sparse, zero inflated, and with a negative skew, so that they are suitable for input to statistical tests without violating the assumptions of the model, and the second is to account for the substantial difference in reads per sample. Normalisation of microbiome (and other types of high throughput sequencing data) represents a considerable methodological challenge and is a statistical research area in its own right.

There is unavoidable irregularity between the proportion and abundance of taxa detected in sample data, compared to the ecosystems from which the samples originate, and this is referred to as "*compositionality*" (Gloor et al. 2017). This is reflected in the variation in number of reads obtained for each sample, generally referred to as "*sequencing depth*" or "*library size*" and must be addressed in addition to the more standard challenges of distribution. A primary aim of normalisation of microbiome data is to approximate the true sequence abundances in the original sample. Secondary aims are to limit false inference and deal with sequence count sparsity. There are numerous methods available for normalisation of the count table, each of which holds relative advantages and disadvantages; no currently available method is without considerable limitations. Further methods are under development.

A widely used method for accounting for sequence depth prior to undertaking downstream analyses is rarefying. This refers to subsetting the sequencing data to a sequencing depth that is common to all samples. For example, in samples within a dataset contain between 10,000 and 100,000 reads, rarefying the samples would remove 90,000 reads from samples containing 100,000, 80,000 from those containing 90,000 and so on, so that all samples are represented by 10,000 reads. Whilst this is arguably the most

effective method possible for dealing with variation in sequencing depth, a prominent criticism of rarefying is that it disregards much data containing useful information. In fact, it exacerbates a general issue inherent to high throughput sequencing techniques; the sequenced sample represents a subset of the ecosystem and is unlikely to be fully representative. This is referred to compositionality – there is an arbitrary total imposed by the instrument (Gloor et al. 2017). This problem can be partially abated by performing multiple rarefying steps on the dataset, for example 500 times, and subsequently taking the average for each ASV or OTU.

Another widely employed method is to calculate relative abundances of taxa. Here, abundances are scaled so that the total abundance of all taxa sums to 1. The theory is that the proportion within the sample should be similar to the proportion in the original sample, and in addition should be directly comparable to taxa abundances in other samples. The assumptions of this technique are that there is a majority random taxon component, and that there is balanced differential abundance between samples; for each number of taxa that increase across samples, there should be the same number of decreasing taxa (Weiss et al. 2017). Calculation of relative abundances results in over inflation of counts of some taxa, with corresponding under-estimation of counts of other taxa. Whilst it is a useful tool for visualisation, other techniques may be preferable for normalisation.

The variance stabilising transformation (VST) which is inherent in DESeq2 and DESeq was originally designed for the analysis of RNAseq data but is now widely employed in microbiome studies. A recent study suggested that the VST provided the most consistent estimation of taxonomic and structural coherence (Badri et al. 2020).

4.4 Beta diversity

Beta diversity metrics quantify the dissimilarity between samples, and as such are measures of the overall compositional difference. Prior to calculation of beta diversity, the data must be normalised – including an adjustment for sequencing depth. This is because an assumption of the model is that samples to be compared are of the same size, either in area or volume (Ricotta and Podani 2017). In this thesis, the count data were normalised using the VST, described above in Section 4.3. Beta diversity was calculated using Bray Curtis, Weighted UniFrac and Unweighted UniFrac metrics.

Bray Curtis is a commonly used non phylogenetic method for measuring beta diversity. It is calculated using two components. Considering two samples, the first component is the abundance of each taxa from sample one minus the abundance of each in sample two, and the second component is the abundance of each taxa in sample one added to the abundance of each taxa in sample two. If a species is absent, its abundance should be recorded as zero. Bray Curtis dissimilarity is then calculated as the sum of the first component divided by the sum of the second component (Bray and Curtis 1957).

UniFrac metrics use phylogenetic information to give an estimate of beta diversity and can be used to detect whether phylogenetic lineages between samples are significantly different. When comparing two samples, Unweighted UniFrac is measured as the fraction of branch lengths in the phylogenetic tree which leads to taxon descendants in one sample or the other, but not both. Weighted UniFrac is similar to Unweighted UniFrac, but with the added component of accounting for differences in the relative abundance of taxa (Lozupone and Knight 2005).

In this thesis, the statistical significance of beta diversity dissimilarity between groups is measured using a permutational multivariate analysis of variance (PERMANOVA) test (Anderson 2001).

4.5 Taxonomic analysis: differential abundance

Microbiota sequence count tables, in which each count represents the number of ASVs or OTUs per sample, are particularly challenging datasets to interpret. However key challenges are common to other comparative high-throughput sequencing assays. These are: small replicate numbers; large dynamic range; discreteness; and presence of outliers. For the purpose of assessing microbiota taxon associations with a phenotype - differential abundance - DESeq2 has become established in the microbiome field. Differential abundance calculation with DESeq2 addresses key challenges for microbiome data and offers a quantitative analysis which focuses on the strength as opposed to presence of differential sequence abundance. DESeq2 uses shrinkage estimation for dispersions and fold changes, to improve stability and reliability of interpretations of estimates. The statistical steps for differential abundance analysis with DESeq2 are as follows.

The DESeq2 model has 3 parameters – (1) a normalisation parameter, (2) a variance parameter and (3) parameters representing group differences. Normalisation in DESeq2 is achieved using the variance stabilising transformation, which goes some way towards addressing variation induced by differences in sequencing depth: some samples will demonstrate higher counts of ASVs simply because more amplicon sequences were yielded for that sample. Application of the variance stabilising transformation yields values, the variances for which are approximately the same throughout the dynamic range within the whole dataset. Fit 2 uses maximum likelihood estimation to estimate sequence-wise dispersion. Dispersion captures the variability between replicates and is critical for statistical inference of differential abundance. After estimation of sequence wide dispersion, a smooth curve is fitted, providing an estimate for the expected dispersion value for sequences but does not represent deviations of sequences from the overall trend. Shrinkage of estimates is then applied, guided by Bayes theorem. The strength of the shrinkage - the extent to which the estimate is moved towards the curve - is dependent

on sample size. In fit 3, DESeq2 employs empirical bayes shrinkage for log fold change estimation (Love et al. 2014a).

4.6 Overview of methods applied in this thesis

Within this thesis an array of approaches are applied to investigate the association between the commensal microbiota and RA. The broad work undertaken within each chapter containing primary analysis is presented in Table 4.1.

	Chapter 3	Chapter 5	Chapter 6	Chapter 7
Microbiome	Gut (stool)	Gut (stool)	Gut (stool)	Oral (saliva)
Stage Started With	Raw sequencing data	ASVs (generated in the work presented in Chapter 3).	Fresh samples collected for project. Received raw sequencing data – used to generate ASVs	ASVs, which were generated by Daniel Sprockett, University of Stanford.
Cohort(s)	TwinsUK (n=3,345)	TwinsUK (n = 1650), SCREEN-RA (n= 133), UK Biobank (n=6,776)	TwinsUK (n=60)	TwinsUK (n=407, with genotyping)
Phenotype	Heritability	Genetic risk of RA	RA	Genetic risk of RA
Measures	ASVs (generation of)	- Host genotype: Polygenic risk score; shared epitope positivity - Microbiome: alpha diversity	- Host genotype: controlled for using RA discordant twin pairs - Microbiome: Alpha diversity, ecological community analysis: balances	- Host genotype: Polygenic risk score - Microbiome: Heritability Alpha diversity, Beta diversity, ASV counts

Statistical Analyses	heritability estimation and consideration of 95% CI.	linear mixed effects model for association of alpha diversity with PRS. Logistic regression for association of PRS with RA diagnosis. ASVs differential abundance against RA PRS.	Microbial differential abundance against RA diagnosis. Linear mixed effects model for association of balances with RA Students T Test; Wilcoxon rank-sum for comparison of alpha diversity	Spearman's correlation of covariates. Microbial differential abundance against RA PRS. Linear mixed-effects model for association of alpha diversity with RA PRS. PERMANOVA for association of beta diversity with RA PRS

Table 4.1 Summary of methods applied
The types of analysis within each chapter containing primary research is summarized.

4.6 Summary

In this chapter I have described the microbiome-specific general methods used within this thesis: alpha diversity, beta diversity and differential abundance. An explanation of other techniques used which are not applicable to the wider thesis are described within the respective chapters: heritability analysis of ASVs is described in Chapter 3, and analysis of the microbial community is described in Chapter 6.

Chapter 5

Genetic risk of rheumatoid arthritis and association with the gut microbiota.

In this chapter I investigate the association between genetic risk of RA and the composition of the gut microbiota, using measures of diversity, taxon abundance association and ecological community interaction.

This chapter is presented as a paper, originally published in The Lancet Rheumatology on the 26th June 2020.

This paper is available online at DOI: 10.1016/s2665-9913(20)30064-3

Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study

Philippa M Wells, Adewale S Adebayo, Ruth C E Bowyer, Maxim B Freidin, Axel Finckh, Till Strowig, Till Robin Lesker, Deshire Alpizar-Rodriguez, Benoit Gilbert, Bruce Kirkham, Andrew P Cope, Claire J Steves, Frances M K Williams**

Lancet Rheumatol 2020; 2: e418–27

See Comment page e375

*Contributed equally

Department of Twin Research and Genetic Epidemiology (P M Wells BSc, A S Adebayo PhD, R C E Bowyer PhD, M B Freidin PhD, C J Steves PhD, Prof F M K Williams PhD), **Centre for Rheumatic Diseases** (Prof A P Cope PhD), **and Centre for Inflammation Biology and Cancer Immunology** (Prof A P Cope), **King's College London, London, UK; Division of Rheumatology, Geneva University Hospital, Geneva, Switzerland** (Prof A Finckh PhD, D Alpizar-Rodriguez PhD, B Gilbert MBBS); **Department of Microbial Immune Regulation, Helmholtz Centre for Infection Research, Braunschweig, Germany** (T Strowig PhD, T R Lesker PhD); **Hannover Medical School, Hannover, Germany** (T Strowig); **Department of Rheumatology, Guy's and St Thomas' NHS Trust, London, UK** (B Kirkham PhD); **and Department of Ageing and Health, St Thomas' Hospital, London, UK** (C J Steves)

Correspondence to:
Prof Frances M K Williams, Department of Twin Research, King's College London, St Thomas' Hospital, London SE1 7EH, UK
frances.williams@kcl.ac.uk

Summary

Background Rheumatoid arthritis is a chronic inflammatory autoimmune disease that is associated with reduced life expectancy. The disease is heritable and an extensive repertoire of genetic variants have been identified. The gut microbiota might represent an environmental risk factor for rheumatoid arthritis. We aimed to assess whether known rheumatoid arthritis risk alleles were associated with the gut microbiota in a large population who do not have rheumatoid arthritis.

Methods In this cross-sectional study done in the UK and Switzerland, we used genotyping and microbiota data from previous studies of the TwinsUK cohort, excluding participants who had ever had a diagnosis of rheumatoid arthritis, as well as their unaffected co-twins. We used blood samples for genotyping and stool samples for the assessment of the gut microbiota. We generated a polygenic risk score (PRS) for rheumatoid arthritis in 1650 TwinsUK participants without the disease, based on 233 GWAS-identified single nucleotide polymorphisms associated with rheumatoid arthritis. We validated the PRS using logistic regression against rheumatoid arthritis diagnosis in 2686 UK Biobank individuals with a confirmed diagnosis of rheumatoid arthritis. Amplicon sequence variants (ASVs) were generated from 16S rRNA gene sequencing of stool samples and assessed for association with the PRS for rheumatoid arthritis. We validated the findings in an independent sample comprised of first-degree relatives of patients with rheumatoid arthritis from the SCREEN-RA cohort. Differential abundance of ASVs present in more than 5% of samples, grouped by ASV taxon annotation, against the rheumatoid arthritis PRS as a continuous variable was assessed using fixed-effects covariates. To account for multiple testing, the false discovery rate calculation was applied to all p values to generate q values, with a significance threshold of 0·05 determined a priori.

Findings We found that presence of *Prevotella* spp were positively associated with the rheumatoid arthritis PRS in TwinsUK participants ($q<1\times10^{-7}$). This finding was validated in SCREEN-RA participants (n=133) carrying established shared epitope risk alleles (q=0·0011). We also found an association between *Prevotella* spp and presence of preclinical rheumatoid arthritis phases (q=0·021).

Interpretation *Prevotella* spp in the gut microbiota are associated with the rheumatoid arthritis genotype in the absence of rheumatoid arthritis, including in individuals at high risk of developing rheumatoid arthritis. Our findings suggest that host genotype is associated with microbiota profile before disease onset.

Funding Versus Arthritis.

Introduction

Rheumatoid arthritis is a debilitating chronic autoimmune condition, associated with reduced life expectancy. The cause of rheumatoid arthritis has a substantial genetic component, with heritability estimated at 65%.[1] Known environmental risk factors include periodontal disease, tobacco smoking, and diet, and these appear to trigger disease onset in genetically susceptible individuals.[2] An additional rheumatoid arthritis risk factors is the mucosal commensal microbiota. Extensive cross-talk exists between microbiota and the host, starting early in life with the development of a normal immune system; microbiota might be implicated in the development of rheumatoid arthritis.

The gut lumen holds most of the commensal microbiota, and has intimate proximity to both the immune system via the gut-associated lymphoid tissue, and the systemic circulation. A key gut microbiota association in patients with rheumatoid arthritis is a relative increase in the abundance of *Prevotella* spp,[3] particularly *Prevotella copri* (*P copri*), which presents early in the course of rheumatoid arthritis.[4,5] *P copri* is also positively associated with clinical parameters in patients with rheumatoid arthritis, further supporting its pathophysiological relevance to the disease.[5] In addition to promoting disease activity, the gut microbiota might also influence the response to treatment of patients with rheumatoid arthritis.[5,6] The gut microbiota therefore represents a potential therapeutic target in

Research in context

Evidence before this study

We searched Google Scholar using the search terms "rheumatoid arthritis microbiome"; "rheumatoid arthritis Prevotella"; "microbiome heritability"; "microbiome, genetics rheumatoid arthritis"; and "rheumatoid arthritis genetic aetiology". We included articles published in English between Jan 1, 1990, and Oct 1, 2019. The reference lists of identified papers were further used to identify relevant literature. *Prevotella copri* has been shown to be increased within the gut microbiota of patients with rheumatoid arthritis, predominantly those with early disease, before treatment is initiated. Prevotellaceae family members have also been shown to be higher in patients with pre-clinical rheumatoid arthritis than in controls. *P copri* is posited to be an inflammatory driver, contributing to rheumatoid arthritis pathology by promoting a pro-inflammatory cytokine milieu.

Added value of this study

This study is, to our knowledge, the first to show that, in a large cohort, carrying genetic risk factors for rheumatoid arthritis is associated with a higher abundance of *Prevotella* spp in the absence of any form of rheumatoid arthritis.

Implications of all the available evidence

Within the gut microbiota, *Prevotella* spp, including but not limited to *P copri*, are of interest in the pathogenesis of rheumatoid arthritis. Our findings suggest that any potential causal role of *Prevotella* spp occurs early in disease development, and therefore targeting of the preclinical stage of rheumatoid arthritis should be explored in future studies.

patients with rheumatoid arthritis, both for the modulation of disease and for improving the response to established therapeutics.

Identifying how genetic and environmental risk factors for rheumatoid arthritis interact with one another might shed light on the underlying biology of the disease. The influence of host genetic factors on the microbiota in rheumatoid arthritis remains somewhat unexplored. An important influence is highly plausible: host genetics shape the biochemical and immune environment in which the microbiota reside, and furthermore the cumulative influence of the genetic risk loci in rheumatoid arthritis is predominantly mediated by immune pathways.[2] Although several factors influence microbiota composition, host genetic factors account for a considerable proportion of variance, with some taxa being 40% heritable.[7]

We aimed to investigate whether genetic risk for rheumatoid arthritis was associated with the composition of the gut microbiota in the absence of clinical disease.

Methods

Study design and participants

This cross-sectional study was done in the UK and Switzerland, and included participants from the TwinsUK (primary) and SCREEN-RA (validation) cohorts. To isolate the genetic influence on disease from the pathophysiology and treatment influence of established disease, TwinsUK participants with rheumatoid arthritis and their co-twins were excluded from the study. We used genotyping and microbiota data from 1650 eligible TwinsUK participants who had contributed blood and stool samples (figure 1). Genetic risk for rheumatoid arthritis in TwinsUK participants was captured using polygenic risk scoring. A polygenic risk score (PRS) for rheumatoid arthritis was generated and validated using logistic regression against rheumatoid arthritis diagnosis in UK Biobank participants, with 2686 confirmed rheumatoid arthritis cases, and applied to rheumatoid arthritis-unaffected participants in TwinsUK. We then calculated the association between the PRS for rheumatoid arthritis and composition of the gut microbiota.

Ethics approval was granted by the St. Thomas' Hospital Research Ethics Committee. Following the restructure and merging of the research ethics committee, subsequent amendments were approved by the National Research Ethics Service (NRES) Committee London–Westminster (TwinsUK reference EC04/015); approval for the use of microbiota samples was granted by the NRES Committee London–Westminster (The Flora Twin Study reference 12/LO/0227).

Participants included in the study were members of the TwinsUK cohort, the largest UK registry of adult twins.[9] We excluded participants who reported ever having had a diagnosis of rheumatoid arthritis, as well as their unaffected co-twin. This strategy allows for isolation of rheumatoid arthritis genetic factors and circumvents the important issues of confounding by rheumatoid arthritis disease and its treatment—thereby achieving a human model of the genetic association with the microbiota in rheumatoid arthritis. The TwinsUK registry is demographically well suited to the study of rheumatoid arthritis (most participants are female and older); as such, we deemed that further study-specific inclusion criteria were not necessary. TwinsUK participants comprised 93% women, with a median age of 63 years (table).

To further confirm the results from TwinsUK, an analysis of genetic predisposition to rheumatoid arthritis and microbiota association was done in participants of SCREEN-RA, a Swiss multi-centre cohort of first-degree relatives of patients with rheumatoid arthritis, who themselves are unaffacted by the disease (figure 1).[8] These relatives share on average 50% of their genotype with their rheumatoid arthritis-affected relative, and thus have a higher genetic risk of rheumatoid arthritis than do those

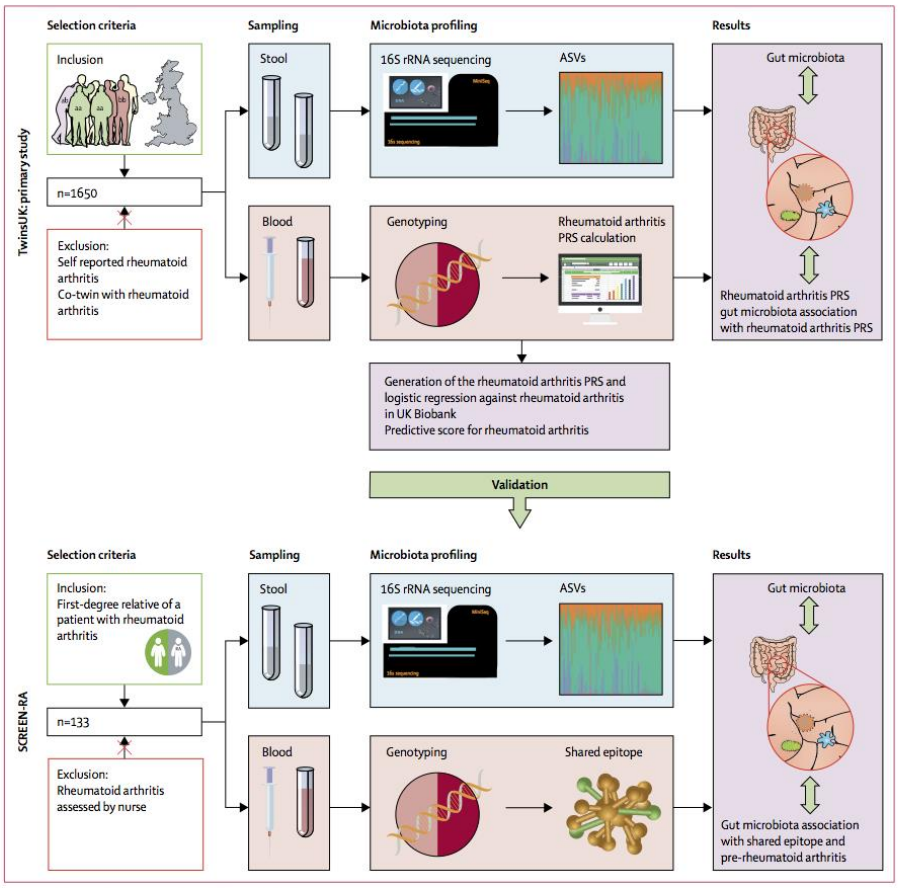


Figure 1: **Schematic representation of the study design**
The sampling, microbiota profiling, and genotyping were previously performed (blue and pink panels).[9] The purple panels indicate analyses done in the current study. The TwinsUK cohort is led by researchers at King's College London, located at St Thomas Hospital NHS Foundation Trust, and comprises adult twins who are resident in the UK. SCREEN-RA is a Swiss multi-centre cohort, which comprises first-degree relatives of patients with rheumatoid arthritis. ASV=amplicon sequence variant. PRS=polygenic risk score.

in the general population. Within SCREEN-RA participants (n=133), the subset of patients with preclinical rheumatoid arthritis (n=83) were identified using the European League Against Rheumatism (EULAR) terminology for preclinical phases of rheumatoid arthritis,[8,10] and matched with 50 controls who were also first-degree relatives of patients with rheumatoid arthritis but who did not have preclinical arthritis. Briefly, preclinical rheumatoid arthritis was defined on the basis of serum positivity for anti-citrullinated protein antibody (ACPA) or rheumatoid factor, or symptoms and signs associated with possible rheumatoid arthritis with or without undifferentiated arthritis.[8,10] Patients with a diagnosis of rheumatoid arthritis were excluded from the analysis and all participants were examined by a trained nurse. SCREEN-RA participants had previously been genotyped for the *HLA-DRB1* shared-epitope risk allele for rheumatoid arthritis, the strongest established genetic association with disease. We examined the gut microbiota composition in relation to the shared-epitope genotype and presence or absence of pre-rheumatoid arthritis in this cohort. Of SCREEN-RA participants, 74 (89%) were women and the median age was 57 years (table). Participants provided written informed consent.

Procedures

Genotyping—TwinsUKcohort

We used sequencing data which had been previously generated as part of the TwinsUK cohort, which is led by the Department of Twin Research and Genetic Epidemiology, King's College London (London, UK). Briefly, blood samples from TwinsUK participants obtained at the clinical visit were used to identify genotype using the Illumina HumanHap300 BeadChip and the Illumina HumanHap610 QuadChip (Illumina, Cambridge, UK). Non-genotyped variants were imputed using 1000 Genomes and Haplotype Reference Consortium reference panels.[9] Testing was undertaken by Affinity Biomarker Labs using the Siemens Centaur XP Anti-Cyclic-C Peptide (Siemens Health Engineers, Frimley, UK). Seropositivity was defined as more than 5 U/mL.

PRS for rheumatoid arthritis in TwinsUK

A PRS to predict rheumatoid arthritis in unaffected TwinsUK participants was calculated and its association determined with composition of the gut microbiota. The PRS assigns an individual a single numerical value for risk of disease conferred by genetic factors.[11] The NCBI database of GWAS summary statistics for rheumatoid arthritis was used to identify 233 published single nucleotide polymorphisms (SNPs) associated with rheumatoid arthritis at genome-wide significance ($p=5\times10^{-8}$), of which 117 had been replicated across studies (appendix p 3).[12] European ancestry was included as a study inclusion criterion, ensuring ethnic concordance with TwinsUK (ie, ethnicities other than white were excluded).[9,13] The PRS was tested for a predictive value for rheumatoid arthritis in 6776 participants from UK Biobank, including 2686 patients with rheumatoid arthritis and 4090 unselected controls with the chronologically closest participant identification numbers. Diagnosis of rheumatoid arthritis in UK Biobank participants had been made using hospital episode statistics data supplied by NHS Digital. All identified rheumatoid arthritis patients were included. We deemed that no exclusion criteria were necessary; the UK Biobank participants and the TwinsUK cohort had similar racial demographics. Logistic regression of 2686 patients with rheumatoid arthritis and unselected controls against PRS, adjusting for age, sex, and smoking history, was applied. Standardised coefficients are reported.

Risk allele dosage of SNPs present within TwinsUK was extracted using PLINK (version 1.9). Of the SNPs identified in the literature, 227 were available in TwinsUK. Pruning was applied to account for linkage disequilibrium.[14] Missing allele dosages were imputed and replaced with the mean value across the respective SNPs. The risk allele dosage was multiplied by the SNP–rheumatoid arthritis association effect size, to produce a weighted PRS.[11]

Microbiota profiling—TwinsUK

We used data that had been previously generated as part of the TwinsUK cohort. Briefly, microbiota composition of stool samples was assessed using the 16S rRNA marker gene, with sequencing of the V4 variable region using barcoded primers (F515/R806). Samples were processed as previously described.[7] Briefly, faecal samples were collected during clinical visits or were posted in sealed ice packs and frozen on arrival at the lab at –80°C. Stool samples were sent as 35 batches on dry ice to Cornell University, NY, USA, where DNA was extracted and sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA).[7]

16S sequences were demultiplexed in QIIME. Amplicon sequence variants (ASVs) were generated using the DADA2 package in R.[15] Sequences were trimmed, error was estimated within the forward and reverse reads for each sample, and the ASV algorithm was applied to infer the original biological sequence. Forward and reverse reads were joined. Chimeras were removed and the total dataset was merged, followed by performing taxonomic assignment using SILVA version 1.3.2.[16] Samples with a sequencing depth of less than 10 000 reads were excluded. A phylogenetic tree was generated using the Phangorn R package. Alpha diversity was calculated on untrimmed ASV tables using four measures: Shannon index, Simpson index, observed ASVs, and Faith's phylogenetic diversity. For the taxonomic analysis, ASVs were grouped according to taxon annotation. Taxonomic assignment using the SILVA database allows for a higher level of differentiation than other databases because in some instances the genus of ASVs can be annotated according to the prediction of species group.[17] These annotations were preserved because they provide more information regarding taxon assignment than genus annotation alone. In this way, *Prevotella*-annotated ASVs might be accurately further

	TwinsUK: rheumatoid arthritis unaffected (n=1650)	SCREEN-RA	
		Pre-rheumatoid arthritis (n=83)	FDR controls (n=50)
Age, years	63 (56–69)	58 (50–66)	55 (47–62)
Sex			
Women	1535 (93%)	74 (89%)	39 (78%)
Men	115 (7%)	9 (11%)	11 (22%)
BMI	25 (23–29)	24 (22–27)	24 (22–27)
ACPA positive	9* (2%)	38 (46%)	0
Rheumatoid factor positive	35*(7%)	28 (34%)	0
Shared epitope positive	..	42 (51%)	32 (64%)
ACPA and rheumatoid factor positive	1 (<1%)	6 (7%)	0
Current smoker	567 (34%)	16 (19%)	11 (22%)
Antibiotic use within the past month	66 (4%)	..	..
Caucasian; northern European	1650 (100%)	83 (100%)	50 (100%)
Swollen joints	..	1 (0–3)	0 (0–1)
Tender joints	..	1 (0–2)	0 (0–1)

Data are median (IQR), or n (%). ACPA positivity was defined as a concentration of greater than 5 uL/mL. Rheumatoid factor positivity was defined as a concentration of greater than 15 uL/mL. FDR=false discovery rate. BMI=body-mass index. ACPA=anti-citrullinated protein antibody. *Serum samples analysed for 500 TwinsUK participants, including those with highest polygenic risk scores.

Table: **Participant demographics**

See Online for appendix

For **QIIME software** see http://qiime.org

For **SILVA software** see https://www.arb-silva.de

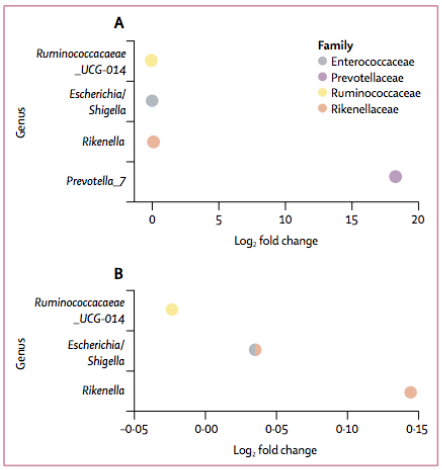


Figure 2: Differential abundance of the gut microbiota per unit increase in rheumatoid arthritis PRS in TwinsUK participants
Positive log-fold change of taxa indicates a positive association with the rheumatoid arthritis PRS. *Prevotella_7* was the strongest association, and positively associated with the rheumatoid arthritis PRS ($q<1\times10^{-7}$). Because of the scale and comparative difference in log-fold change in *Prevotella*, this taxon was excluded (B) to allow visualisation of the three other associations. Other taxa associations within the gut microbiota were *Ruminococcaceae_UCG-14* (q=0·045), *Rikenella* (q=0·018), and *Shigella* (q=0·018). PRS=polygenic risk score.

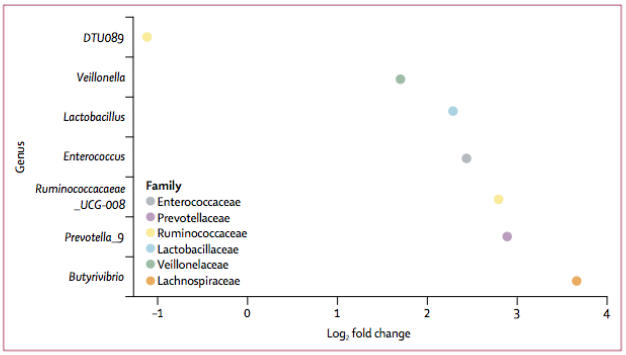


Figure 3: Differential abundance of the gut microbiota in patients with preclinical rheumatoid arthritis compared with unaffected controls in the SCREEN-RA cohort
Prevotella_9 was associated with preclinical rheumatoid arthritis (q=0·021). We found six additional genus group associations: *Lactobacillus* (q=0·0003), *Butyrivibrio* (q=0·018), *Ruminococcaceae_UCG-D08* (q=0·018), *Enterococcus* (q=0·018), *DTU089* (q=0·042), and *Veilllonella* (q=0·042).

differentiated using 16S data, which has been a methodological challenge previously.

ACPA testing–TwinsUK

500 TwinsUK participants, including 250 (50%) participants with high genetic risk (PRS within the upper quartile) were tested for ACPA seropositivity. Testing was done by Affinity Biomarker Labs, using Siemens Centaur XP Anti-cyclic-C peptide kit (Siemens Health Engineers, Frimley, UK). ACPA seropositivity was defined as more than 5 U/mL.

Phylogenetic and community relationship within the Prevotellaceae family

To investigate the *Prevotella* spp associations that were identified further, the phylogenetic and community relationships between the two implicated groups of ASVs (*Prevotella_7* and *Prevotella_9*) were explored. For these methods, grouping ASVs by taxon annotation was not appropriate and ASVs were considered as singular units. To examine the microbial ecological community relationships, ASVs were grouped into compositional clusters, or balances. Briefly, ASVs that were present in at least 10% of individuals were correlated and transformed as implemented by Morton and colleagues.[18] This method creates clades in which interacting species are closer neighbours in a clade than loosely related ones are.

Shared epitope genotyping—SCREEN-RA cohort

As a validation step, we examined the gut microbiota composition in first-degree relatives of rheumatoid arthritis patients (SCREEN-RA cohort) in relation to shared epitope positivity, and a pre-disease state as defined by clinical rheumatoid arthritis parameters.[8,10]

We used sequencing data which had been previously generated as part of a previous study of the SCREEN-RA cohort. We contacted the corresponding author of Alpizar-Rodriguez and colleagues' study[8] who co-leads the SCREEN-RA cohort, and they agreed to share their SCREEN-RA cohort data. Briefly, Alpizar-Rodriguez and colleagues[8] obtained blood from SCREEN-RA participants and controls. DNA was extracted using a modification of the salt-out technique (Nucleon TM, Scotlab, Glasgow, UK). *HLA-DRB1* shared epitope polymorphism was assessed by reverse PCR (sequence-specific oligonucleotide primers hybridisation and by PCR) sequence-specific primers (SSP) using commercial reagents validated at the Swiss National Reference Laboratory for Histocompatibility. The method discriminates all major subtypes in different allele groups within the *DRB1*04* genotype. *HLA-DR1, DR4*, and *DR14* alleles that were negative for the shared epitope 70–74 motif were also discriminated. In a second step, the shared epitope-positive typing ambiguities were analysed by PCR-SSP to find the final 4-digit typing result.

Microbiota profiling—SCREEN-RA cohort

Gut microbiota 16S rRNA data was collected as previously described.[8] Briefly, the DNA Genotek OMNIgene Gut Stool Microbiome Kit (DNA Genotek, Ottawa, ON, Canada) was used to collect, store, and ship the stool samples. After sample processing and DNA extraction, the V4 region of the 16S rRNA gene was amplified using barcoded primers (F515/R806), and sequenced on an Illumina MiSeq

(Illumina, San Diego, CA, USA), with ASVs generated as per the TwinsUK cohort to ensure compatibility.

Statistical analysis

Linear mixed-effects models were used to determine association between rheumatoid arthritis PRS and alpha diversity, using alpha diversity as a response variable to the rheumatoid arthritis PRS as a continuous variable. Modelling was performed using the lme4 package in R,[19] with fixed effect covariates age, body-mass index, and sequencing depth and technical covariates as random effects. Standardised coefficients were reported.

Differential abundance of ASVs present in more than 5% of samples, grouped by ASV taxon annotation, against the rheumatoid arthritis PRS as a continuous variable was assessed using the DESeq2 R package,[20] using fixed-effects covariates. To account for multiple testing, the false discovery rate calculation was applied to all p values to generate q values, with a significance threshold of 0·05 determined a priori.

To examine the phylogenetic relationship, the phylogenetic tree was subsetted to Prevotellaceae and visualised using the Phyloseq R package.[21]

Differential abundance of taxa in the gut microbiota in association with preclinical rheumatoid arthritis and shared epitope positivity was assessed against all genus present in greater than 5% of samples using the DESeq2 R package.[20] Biological covariates were not required to adjust for because these factors were not statistically different between the case and control groups (appendix p 1). We used the same method as Alpizaar-Rodriguez and colleagues,[8] with the advancement that ASVs and DESeq2 methods were used.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. PMW, ASA, MBF, RCEB, CJS, and FMKW had access to the raw data. The corresponding authors had full access to all of the data and the final responsibility to submit for publication.

Results

1650 TwinsUK participants without rheumatoid arthritis were included, most of whom were women (table). The PRS was normally distributed in the TwinsUK sample with patients with rheumatoid arthritis and their twin siblings excluded (Shapiro-Wilk normality test p=0·24). Logistic regression of the PRS in 6776 UK Biobank participants, of which 2686 (39·6%) had a diagnosis of rheumatoid arthritis according to hospital episode statistics, confirmed that the score is predictive of rheumatoid arthritis (odds ratio per SD 1·34; p=4·17×10^{-8}).

Individuals with a high PRS in TwinsUK did not show rheumatoid arthritis seropositivity: only nine (2%) of 500 individuals in the TwinsUK sample who had serum available were positive for ACPA, defined as more than 5 U/mL (table), and were distributed similarly across genetic risk groups, by PRS quartile (high 2:93; low 7:398).

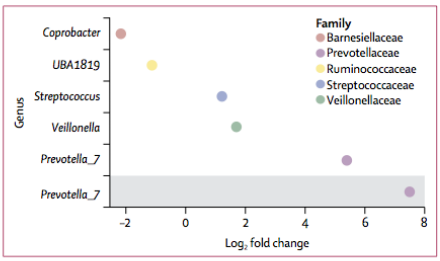


Figure 4: **Differential abundance of the gut microbiota in *HLA-DRB1* shared epitope-positive participants, compared with shared epitope-negative controls in the SCREEN-RA cohort**
In all participants, *Prevotella_7* showed the most substantial positive log-fold change of 5 (q=0·0348). Other genuses that showed lower but significant positive log-fold associations were *Veillonella* (q=0·035), *Streptococcus* (q=0·035), *Ruminococcaceae UBA1819* (q=0·035), and *Coprobacter* (q=0·0020). In participants without symptoms associated with rheumatoid arthritis (excluding tender or swollen joints; grey shading), *Prevotella_7* solely remained positively associated with shared epitope positivity (q=0·0011).

We investigated whether genetic risk of rheumatoid arthritis was associated with alpha (within sample) diversity. We found no detectable association between PRS for rheumatoid arthritis and Shannon index (p=0·76), Simpson index (p=0·41), observed ASVs (p=0·60) or Faith's phylogenetic diversity (p=0·65; appendix p 2). Alpha diversity measures indicate microbial density (observed ASVs) and distribution, in which the distribution of more abundant versus less abundant taxa is assessed, or both of these factors (Shannon index, Simpson index). Faith's phylogenetic diversity is based on the phylogenetic distance between taxa in a sample.

The gut microbiota were taxonomically assessed for association with PRS for rheumatoid arthritis following a non-targeted approach. *Prevotella* ASVs were grouped together according to predicted species and given a numerical designation in the SILVA database. Within these groups, *Prevotella_9* was predicted to be *P copri*, whereas *Prevotella_7* was annotated to multiple *Prevotella* spp, with low sequence divergence. Of all 172 microbial taxa that were present in the gut (stool) microbiota of more than 5% of participants, *Prevotella_7* had the strongest taxon association with the PRS for rheumatoid arthritis (18-fold log base 2 higher differential abundance; q<1×10^{-7}; figure 2). No additional *Prevotella* assocations with the PRS for rheumatoid arthritis were found in the TwinsUK cohort.

We sought to confirm our TwinsUK findings in the SCREEN-RA cohort by analysing the association between the main genetic risk factor for rheumatoid arthritis—the HLA-DRB1 shared epitope—and the gut microbiota. We also investigated the association of the gut microbiota with preclincal stages of rheumatoid arthritis in SCREEN-RA participants.

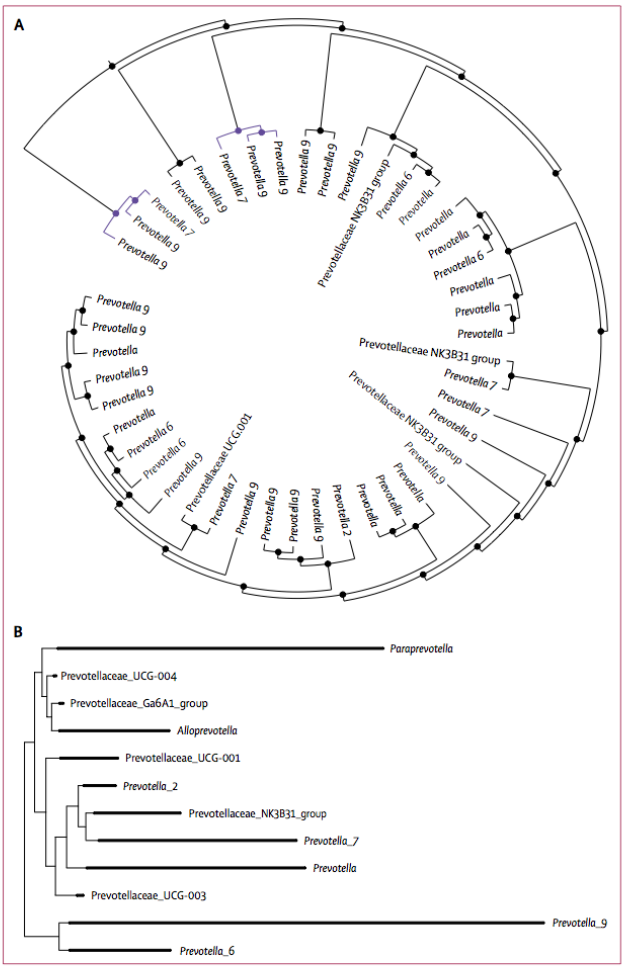


Figure 5: **Relationship between *Prevotella* ASVs within stool samples from TwinsUK participants**
(A) Cluster tree showing an ecological community relationship between *Prevotella* ASVs within stool samples from TwinsUK participants. Taxa clades represent members of the same microbial ecological community. *Prevotella_9* and *Prevotella_7* are the only two taxon clusters present, and ASVs assigned to these taxa cluster in the same clade more frequently than do those in any other taxa. Clustering of *Prevotella_7* and *Prevotella_9* ASVs (shown in purple) indicates that the abundance of these taxa is interdependent; a biological interdependence and environmental niche similarity is suggested. Incremental levels are shown, from individual ASVs joined as two-taxon clusters, to the full set of Prevotellaceae ASVs joined by the final head node. (B) Phylogenetic tree for Prevotellaceae ASVs within TwinsUK participants. *Prevotella_9* and *Prevotella_7* are phylogenetically distinct from each other. ASV=amplicon sequence variant.

In the SCREEN-RA cohort, *Prevotella_9* was positively associated with preclinical rheumatoid arthritis (q=0·021; figure 3). *Prevotella_7* was associated with HLA-DRB1 shared-epitope risk alleles for rheumatoid arthritis in the SCREEN-RA cohort (n=133; q=0·035); the association was stronger in a subgroup analysis in which 44 participants with swollen joints were removed to isolate genotype from rheumatoid arthritis pathophysiology (n=89; q=0·0011; figure 4). In the subgroup analysis of differential abundance against shared-epitope positivity in asymptomatic participants only, *Prevotella_7* was the only remaining taxon association. The key implication is that *Prevotella_9* and *Prevotella_7* are distinct from one another, and *Prevotella_9* is predicted to be *P copri*.[16,19]

As *Prevotella* spp were the strongest taxon associations in our results and of particular interest in rheumatoid arthritis, we further investigated the phylogenetic and community relationships within Prevotellaceae. In particular, we were interested in the relationship between *Prevotella_7* (which we identified to be associated with genetic risk) and *Prevotella_9* (which we identified to be associated with preclinical rheumatoid arthritis). Cluster analysis showed a community relationship between *Prevotella* spp. *Prevotella_7* and *Prevotella_9* were the only two species clusters identified within the Prevotellaceae family, and they clustered together more frequently than any other group did, suggesting an interdependent community relationship between these taxa (figure 5). Visualisation of the Prevotellaceae phylogenetic tree showed that *Prevotella_7* and *Prevotella_9* are phylogenetically distinct from one another (figure 5B).

Prevotella_7 might be implicated in rheumatoid arthritis pathogenesis, and might be increased in patients with rheumatoid arthritis (appendix p 1). Conversely, given that the rheumatoid arthritis-unaffected TwinsUK participants were beyond the mean age of rheumatoid arthritis onset, yet some had high genetic risk scores, these results might hypothetically reflect participants being resilient to the development of rheumatoid arthritis. Correspondingly, the genetic risk and microbiota associations showed might potentially be microbial markers of rheumatoid arthritis resilience. In this instance, the ratio of *Prevotella_7* to *Prevotella_9 (P copri)* would be expected to be higher in unaffected twins compared with rheumatoid arthritis-affected twins.

The relative abundance of *Prevotella* spp was calculated in 18 monozygotic female rheumatoid arthritis-discordant twin pairs who were excluded from our previous analysis, in whom rheumatoid arthritis diagnosis had been confirmed during clinical visit. A higher relative abundance of *Prevotella_7* was found in the rheumatoid arthritis-affected twins compared with unaffected control twins (relative abundance 0·005 *vs* 0·001; appendix p 2). This finding neither supports nor rejects the hypothesis that *Prevotella_7* is implicated in rheumatoid arthritis aetiology. Abundance of *Prevotella* spp was not significantly associated with rheumatoid arthritis. The proportion of *Prevotella_7* (species prediction uncertain) to *Prevotella_9* (predicted *Prevotella copri*) was higher in the rheumatoid arthritis-affected twins than in the control twins (0·4 *vs* 0·01), suggesting that

Prevotella_7 is not a marker of rheumatoid arthritis resilience.

Discussion

Our results show a link between host genetic risk for rheumatoid arthritis and the gut microbiota in two large cohorts. We found associations between gut microbiota and genetic risk in the absence of clinically detectable disease in TwinsUK. The strongest association between rheumatoid arthritis genetic risk and the gut microbiota was an increase in abundance of *Prevotella_7*. This finding was validated in the SCREEN-RA cohort, in which *Prevotella_7* was positively correlated with shared-epitope positivity. *Prevotella_7* was shown to share a biological interdependence with *Prevotella_9* (predicted *P copri*).

Considerable interest exists in *P copri* as a potential mediator of rheumatoid arthritis pathology, and the bacteria is a candidate keystone taxon enriched in the gut microbiota of patients with newly diagnosed rheumatoid arthritis. Since this observation was made, the human immune response to this microbe has been of considerable interest. Antibodies to *P copri* have been shown to associate with disease severity and T helper type 1 (Th1) cell-mediated and Th17 cell-mediated immune responses in patients with rheumatoid arthritis.[22] Functional work has suggested a role for *P copri* in Th17 cell differentiation.[23] That *P copri* is associated with new-onset rheumatoid arthritis before treatment with disease-modifying antirheumatic drugs,[4] and is also associated with other inflammatory conditions,[24] has fuelled speculation that inflammation is a prerequisite for *P copri* proliferation within the gut, relative to other taxa.[22,25] *P copri* might have adapted to thrive in a pro-inflammatory environment, and might further promote the inflammatory milieu, thereby enhancing the bacteria's own favoured environmental niche.[25] In doing so, *P copri* is suggested to contribute to rheumatoid arthritis pathology. According to this model, a human–microbiota interspecies positive feedback loop is proposed. Another example of such a model is the hijacking of complement cascade by *Porphyromonas gingivalis*.[26]

An increase in abundance of three other taxa (*Ruminococcaceae_UCG-14*, *Rikenella*, and *Shigella*) with the PRS for rheumatoid arthritis was also observed. Ruminococcaceae and *Shigella* have been shown previously to be in higher abundance in rheumatoid arthritis patients compared with controls.[3,4] The evidence for a role of these taxa in rheumatoid arthritis is much weaker than for *P copri*, with lower replication across studies, and no functional link reported to date. However, the association both with patients with rheumatoid arthritis and with genetic risk for rheumatoid arthritis in a large, cohort without disease is interesting, and merits further investigation.

Although causality is not knowable from a cross-sectional association study, our results provide robust evidence indicating that host genetic factors influence the abundance of *Prevotella* in the gut micobiota. Both TwinsUK and SCREEN RA cohorts were balanced for ACPA positivity in relation to rheumatoid arthritis genetic risk loci. A previous study of patients with new-onset rheumatoid arthritis and healthy participants reported an inverse association of HLA genotype with abundance of *P copri*, in the opposite direction to that expected.[5] This association might potentially relate to population differences or confounding factors that were not examined. The present study, using ASVs in a large sample without disease, provides more substantial evidence for an association of genotype with the rheumatoid arthritis gut microbiota.

Our findings suggest that the microbiota are altered before disease, which is in accordance with previous reports of *Prevotella* spp in patients with pre-rheumatoid arthritis[8] and rheumatoid arthritis.[5] Indeed, we found similar results in the patients with preclinical rheumatoid arthritis in first-degree relatives of patients with rheumatoid arthritis. In a study of the SCREEN-RA cohort,[8] an enrichment of Prevotellaceae in the gut microbiota in patients with pre-rheumatoid arthritis was observed. In that study,[8] no specific Prevotellaceae genera were shown to drive the association. Therefore, in this study we reanalysed these data using the more recent method of ASVs. ASVs offer an updated approach to traditional clustering based methods generating operational taxonomic unit (OTUs) from 16S sequences. As opposed to grouping sequences on the basis of a similarity threshold as for OTUs, the error rate is used to infer the original biological sequence and produce units of matched sequence. Therefore, ASVs offer higher resolution and have greater biological relevance than OTUs.[27] The ASV analysis showed a false discovery rate-adjusted significant *Prevotella* genus association with a species-level group annotation, indicating potential *P copri*. Our follow-up investigation of patients with pre-rheumatoid arthritis revealed a further six novel genus associations, which had not been evident in the original OTUs—*Lactobacillus*, *Enterococcus*, *Faecalibacterium*, Ruminococcaceae UBA1819, *Veillonella*, and *Butyrivibrio*. Of the six associations, the first four have been reported to be associated with rheumatoid arthritis,[4,28,29] and Ruminococcaceae were associated with rheumatoid arthritis PRS in the TwinsUK cohort. *Butyrivibrio* has not yet been associated with rheumatoid arthritis, but is a physiologically interesting taxa associated with the production of short chain fatty acid and host metabolism.

This study has several limitations. First, in the pre-rheumatoid arthritis follow-up analysis both cases and controls were first degree relatives of patients with rheumatoid arthritis, with increased genetic risk compared with the general population. Patients with preclinical rheumatoid arthritis could have had higher genetic risk than the controls did, but because full genotyping was not available, we were unable to confirm this possibility. The TwinsUK cohort is predominantly female. The SCREEN-RA cohort is slightly more balanced in terms of sex; however, population prevalence of rheumatoid arthritis is much higher in women than men. Because the

age of onset of rheumatoid arthritis is 30–65 years,[30] the TwinsUK participants, who have a median age of 63 years, are less likely to develop disease than are the SCREEN-RA participants. However, the age of the cohort is of benefit to the design of our study of genetically high-risk yet unaffected individuals and helps us to understand microbiota differences in the absence of disease.

This study highlights the value of using the newer methods of ASVs, which can detect taxonomic variation overlooked by OTU-based methods. Further microbiota associations with rheumatoid arthritis are to be anticipated from improvements in sequencing and interpretation. Annotation of ASVs with the SILVA database[16,17] showed differences at the species level, which would be overlooked using other reference databases. Finally, modelling of the community relationship provides valuable insight into the underlying biology. Future studies should take advantage of these methods.

Taken together, these results support the hypothesis that microbiota is altered in individuals with genetic predisposition to rheumatoid arthritis before the onset even of preclinical rheumatoid arthritis. Our findings shed light on our understanding of microbiota in rheumatoid arthritis and addresses the issue of cause versus consequence—if microbial alteration precedes disease, the microbiota might lie on the causal pathway. However we cannot yet exclude the possibility that *Prevotella* spp are bystanders. Additionally, having genetic risk for rheumatoid arthritis and *Prevotella* spp is likely not sufficient for disease development, but rather might be one of several insults contributing to progression of rheumatoid arthritis pathology, in line with the favoured two-hit hypothesis of rheumatoid arthritis pathogenesis.[31] The identification of preclinical rheumatoid arthritis represents an important clinical target in early disease intervention and is the subject of multiple immune-modulating clinical trials. Further, the genetic risk–microbiota associations that we identified might be applicable to other diseases because a crossover exists in genetic cause between rheumatoid arthritis and other autoimmune conditions. Finally, the microbiota might offer the opportunity for modulation of pre-disease pathways alone or in combination with immune-modulating drugs.

In conclusion, gut microbiota abundance is associated with the genotype for rheumatoid arthritis risk even in the absence of disease. Genotype might mediate key taxonomic associations of the gut microbiota with rheumatoid arthritis, particularly *Prevotella* spp,[4,5,22,23] suggesting that these species play a role early in the development of rheumatoid arthritis.

Contributors

CJS, FMKW, BK, and APC conceived the project. PMW developed the theory, did the experiments, analysed and interpreted the data, and took the lead in writing the manuscript. CJS and FMKW supervised the analysis. CJS encouraged development of the work through leading collaboration with the SCREEN-RA cohort, and inclusion of investigation of community structure, and abundance of taxa in rheumatoid arthritis discordant twins. AF, TS, TRL, DA-R, and BG shared data generated from SCREEN-RA cohort and encouraged the investigation of rheumatoid arthritis autoantibodies in the TwinsUK cohort. ASA did the analysis of the community structure. FMKW, PMW, and CJS collected data from TwinsUK. RCEB contributed to the design and interpretation of statistical models. MBF assisted with UKBiobank genotype data. All authors provided ctitical feedback and contributed to the final manuscript.

Declaration of interests

We declare no competing interests.

Data sharing

Data used in this study are available upon reasonable request to TwinsUK.

Acknowledgments

This work was funded by a Versus Arthritis Special Strategic Award (grant 21227). Twins UK receives funding from the Wellcome Trust (including phenotypic and genotypic data under WT081878MA); the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. CJS receives funding from the UK Medical Research Council, Wellcome Trust, and Clinical Disease Reasearch Foundation. PMW is supported by a KMRT PhD studentship from King's College London. The funding providers for this work did not provide input regarding study design. The authors would like to thank the participants of the TwinsUK and SCREEN-RA cohorts. We thank Marina Mora-Ortiz and Louca Panayioutis for assisting with the figures.

References

1 MacGregor AJ, Snieder H, Rigby AS, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000; **43:** 30–37.
2 Wells PM, Williams FMK, Matey-Hernandez ML, Menni C, Steves CJ. RA and the microbiome: do host genetic factors provide the link? *J Autoimmun* 2019; **99:** 104–15.
3 Sun Y, Chen Q, Lin P, et al. Characteristics of gut microbiota in patients with rheumatoid arthritis in Shanghai, China. *Front Cell Infect Microbiol* 2019; **9:** 369.
4 Zhang X, Zhang D, Jia H, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 2015; **21:** 895–905.
5 Scher JU, Sczesnak A, Longman RS, et al. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* 2013; **2:** e01202.
6 Isaac S, Artacho A, Nayak R, et al. Op0119 the pre-treatment gut microbiome predicts early response to rheumatoid arthritis therapy. *Ann Rheum Dis* 2019; **78:** 133–34.
7 Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. *Cell* 2014; **159:** 789–99.
8 Alpizar-Rodriguez D, Lesker TR, Gronow A, et al. *Prevotella copri* in individuals at risk for rheumatoid arthritis. *Ann Rheum Dis* 2019; **78:** 590–93.
9 Verdi S, Abbasian G, Bowyer RCE, et al. TwinsUK: the UK adult twin registry update. *Twin Res Hum Genet* 2019; **22:** 523–52.
10 Gerlag DM, Raza K, van Baarsen LG, et al. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. *Ann Rheum Dis* 2012; **71:** 638–41.
11 Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018; **50:** 1219–24.
12 Jones HJ, Hubbard L, Mitchell RE, et al. Association of genetic risk for rheumatoid arthritis with cognitive and psychiatric phenotypes across childhood and adolescence. *JAMA Netw Open* 2019; **2:** e196118
13 Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* 2019; **10:** 1–9.
14 Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* 2015; **97:** 576–92.
15 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13:** 581–83.

16 Callahan B. Silva taxonomic training data formatted for DADA2 (Silva version 132). *Zenodo* 2018; published online Feb 13. DOI:10.5281/zenodo.1172783.

17 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15:** 550.

18 Morton JT, Sanders J, Quinn RA, et al. Balance trees reveal microbial niche differentiation. *mSystems* 2017; **2:** e00162-16.

19 Henderson G, Yilmaz P, Kumar S, et al. Improved taxonomic assignment of rumen bacterial 16S rRNA sequences using a revised SILVA taxonomic framework. *PeerJ* 2019; **7:** e6496.

20 Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015; **67:** 1–48.

21 McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013; **8:** e61217.

22 Pianta A, Arvikar S, Strle K, et al. Evidence for immune relevance of *Prevotella copri*, a gut microbe, in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2017; **69:** 964–75.

23 Maeda Y, Kurakawa T, Umemoto E, et al. Dysbiosis contributes to arthritis development via activation of autoreactive T Cells in the intestine. *Arthritis Rheumatol* 2016; **68:** 2646–61.

24 Pedersen HK, Gudmundsdottir V, Nielsen HB, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 2016; **535:** 376–81.

25 Hofer U. Microbiome: pro-inflammatory *Prevotella*? *Nat Rev Microbiol* 2014; **12:** 5.

26 Olsen I, Lambris JD, Hajishengallis G. Porphyromonas gingivalis disturbs host–commensal homeostasis by changing complement function. *J Oral Microbiol* 2017; **9:** 1340085.

27 Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* 2017; **11:** 2639–43.

28 Picchianti-Diamanti A, Panebianco C, Salemi S, et al. Analysis of gut microbiota in rheumatoid arthritis patients: disease-related dysbiosis and modifications induced by etanercept. *Int J Mol Sci* 2018; **19:** e2938.

29 Chen J, Wright K, Davis JM, et al. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med* 2016; **8:** 43.

30 Mueller RB, Kaegi T, Finckh A, Haile SR, Schulze-Koops H, von Kempis J. Is radiographic progression of late-onset rheumatoid arthritis different from young-onset rheumatoid arthritis? Results from the Swiss prospective observational cohort. *Rheumatology* 2014; **53:** 671–77.

31 McInnes IB, Schett G. Pathogenetic insights from the treatment of rheumatoid arthritis. *Lancet* 2017; **389:** 2328–37.

Chapter 6

Analysis of the gut microbiota in rheumatoid arthritis discordant twin pairs

In the prior chapter, I focus on the association of the RA gut microbiota with host genotype. In this chapter, I take the opposite approach, considering the differences in the intestinal microbiota whilst taking account of host genotype. To achieve this, I undertake comparison of the intestinal microbiota of RA affected participants of TwinsUK, with their RA unaffected twin siblings.

Collaborator Attributions

Flore Zuffery created a subset of RA discordant twins for a prior study and carefully phenotyped them; these participants formed a subset of the present study, sample defined below. In this process, Catherine Redshaw, Rachel Horsfall, Quincy D'Souza and I recruited twins using an algorithm to attend study visit, including sampling of their faecal microbiome. Ruth Bowyer had analysed the dietary data for TwinsUK to generate the Healthy Eating Index (HEI).

6.1 Introduction

RA is associated with compositional differences in the intestinal microbiota with potential aetiological relevance, however it is unclear which factors may underlie the variation observed; there is an interest, sparked by encouraging findings which support a role of the intestinal microbiota in early RA in promoting inflammation generated over the past decade (Scher et al. 2013, Zhang et al. 2015, Maeda et al. 2016). This work has focused on *P.copri* and T cell differentiation, though the picture is far more complex (reviewed in Chapter 2). Inspired by these findings, there is considerable research effort to understand the microbiota and its potential pathogenic relevance during the several different stages of RA development and its treatment. Two prior studies of the intestinal microbiome in RA have included an established (treated) RA group (Scher et al. 2013, Zhang et al. 2015). Both studies indicated that the microbiota is different in established (treated) RA versus new-onset (untreated RA), while the intestinal microbiota of new-onset RA individuals was more similar to healthy controls than that of established RA patients (Zhang et al. 2015). This pattern of treatment of a condition resulting in a "normalisation" of the microbiota is quite unusual, and merits further investigation. Overall, the study of the microbiome in RA is in the exploratory phase. This is particularly true with regards to established (treated) RA, as prior work has largely focused on early, untreated RA (Scher et al. 2013, Zhang et al. 2015, Kim and Kim 2016, Larsen 2017).

Moving forward, studies are required to further characterise and provide clarification of the differences in the RA intestinal microbiota compared to healthy controls. In doing so, it will be important to understand which factors may be driving these differences – which constituents of the RA microbiota are driven by modifiable factors, and which are driven by intrinsic or early-life factors. This will provide insight into the aetiological process. It will also be important to understand if there are robust differences in the RA microbiota, whether these aggravate the pathology, and if so what are the potential means of modifying the microbiota with respect to the pathology of RA. As part of this, it will be

informative to derive which differences are associated with the disease state itself, and which are associated with other influences. Picking apart this information constitutes a challenge.

In order to understand the association of the microbiota with the RA diseased state, it is necessary to consider confounding factors. There are multiple factors which influence the composition of the commensal microbiota, including age, sex, BMI, diet, bowel habit, genetics, xenobiotics, birth mode and other early life factors, and geographical environment (discussed in Chapter 1). Most studies of the human intestinal microbiota include statistical consideration of age, sex and BMI within the study design. However other important influences are almost impossible to control for within a standard case and control study: host genotype, and early life factors such as birth mode, childhood diet, and childhood domestic and geographical environment.

Disease discordant twins – where just one of a twin pair is affected by the disease of interest - share crucial physiological influences: genetic factors, pre-natal environment, birth, and childhood environmental and lifestyle factors, but diverge on disease phenotype. The use of same sex twin pairs therefore provides a powerful means to account for confounding factors, which limit classical case-control studies.

A challenge regarding identification of differentially abundant taxa in relation to a phenotype is that taxon abundances are biologically interdependent. That is, the abundance of one taxon may influence the abundance of another or could flourish in related niches without directly influencing one another. Whilst investigation of individual taxon associations has advantages, such as the detection of lower abundance of differentially abundant taxa, there are key underlying biological tenets that are largely ignored using these approaches; bacteria do not exist as single entities but as communities. The co-abundance of bacteria may be understood from an ecological perspective, with competition for resources, adaptation to the environment and interdependence as prominent driving factors. Broadly, microbial interactions can be classified as parasitism,

competition, predation, or mutualism (Morton et al. 2017). Where the environment allows, bacteria form intricate community structures known as biofilms. Within biofilms, bacteria communicate with and influence each other via quorum sensing and outer membrane vesicles (OMVs), to form a specialised community, with different functional areas within the biofilm, encapsulated by a protective polysaccharide layer. The healthy intestinal tract provides a hostile environment for the formation of biofilms having a thick mucosal layer and higher motility. However high-turnover biofilm formation may occur within the colon – some taxa form mucin adhesive pili, which serve as anchors, and polysaccharides, which serve to stabilise the microbial environment (Macfarlane et al. 2011, Vos 2015).

In this chapter I investigate the association of the gut microbiota with RA, using RA discordant TwinsUK twin pairs. This is a novel approach in understanding the RA microbiome. I investigate the association of microbiota diversity and taxa abundance with RA. Further, I investigate the intestinal microbiota ecological community differences between RA affected and control twins.

6.2 Methods

6.2.1 Study Participants

Participants in the TwinsUK database who are likely to have RA, based on self-report measures, were identified using responses to self-report questionnaires. Participants were asked as part of a wider survey if they had been diagnosed with RA by a doctor. This was asked repeatedly to participants in 7 questionnaires between the years 1999 and 2014. Participants were also asked using a self-report questionnaire whether they had been prescribed medication for RA, and if so to specify the medications. This was asked during visits to the clinical research facility between the years 2004 and 2014. Response rates and consistency of responses varied considerably. In addition, self-report questionnaires are

limited in accuracy. For example, participants may be confused as to which type of arthritis they have been diagnosed with, with osteoarthritis being much more prevalent across all age groups except the youngest. This presented a challenge in determining which participants within the database are affected by RA. To address this problem, I collated responses to RA diagnosis questions, and calculated the sum of positive responses for each participant. Within the medication data, participants who had been prescribed disease modifying antirheumatic drugs (DMARDs) and biologic therapies, which have a high specificity for RA, were identified using a fuzzy matching method; this approach identifies words within a large dataset, in which words (in this case drug names), may be misspelt by identifying similar groups of characters. After cleaning the medication data, the medication and diagnosis data were collated. Participants who had responded positively to the most recent RA diagnosis question, participants who had given at least three repeated positive responses to RA diagnosis questions, and participants who had given at least one positive response to RA diagnosis questions and who had also reported prescription of DMARDs or biologics for RA, were listed as potential RA cases within the dataset. This approach was used to generate lists of potential RA discordant twin pairs within the TwinsUK dataset. Participants were then contacted by phone by a research assistant or myself, and asked to confirm whether they had been diagnosed with RA. Those who confirmed an RA diagnosis, and their co-twin (who also confirmed that they had not been diagnosed with RA), were invited to attend a clinical visit in which RA diagnosis was confirmed by a rheumatologist or trained research assistant, and stool and blood samples taken for rheumatoid factor and anti-CCP antibodies. Postal samples were used for one of the RA affected participants (previously identified by Flore Zuffery) and two of the control participants as they were unable to attend the clinical research facility in person. The protocol for study recruitment is summarized in Figure 6.1.

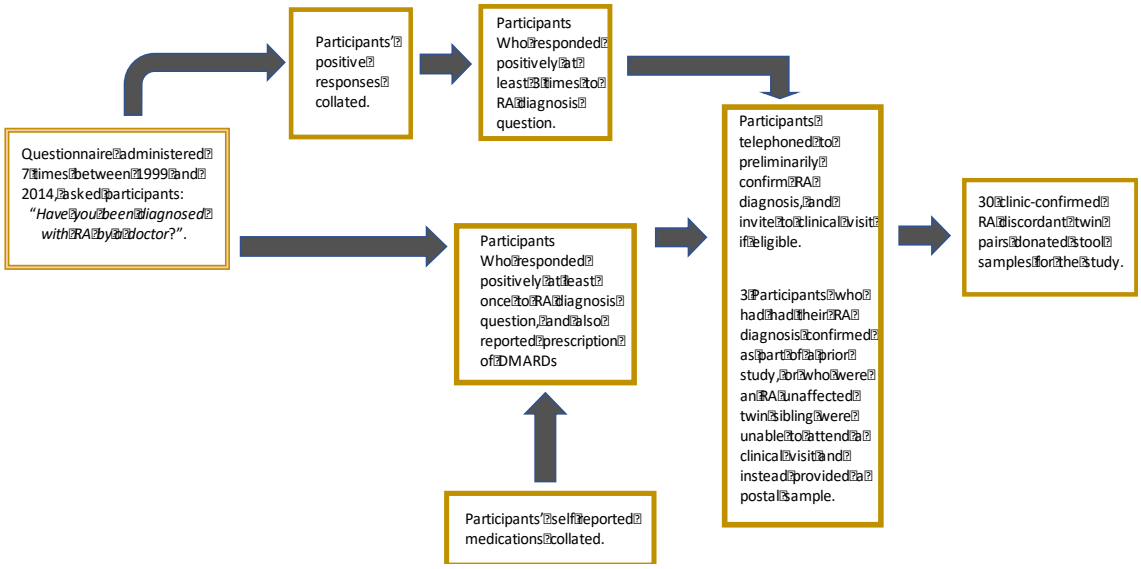


Figure 6.1 Flow diagram of study recruitment protocol.

Participants were predominantly female (90%) and entirely Caucasian, with a median age of 65. Participant demographics are summarized in Table 6.1.

	Group	Age Median (IQR)	Sex (%F)	BMI Median (IQR)	HEI Median (IQR)	Ethnicity
MZ & DZ (n=60)	RA (n=30)	65 (16)	90	26.0 (4.2)	..	Caucasian
	Control (n=30)	65 (16)	90	27.1 (4.31)	..	Caucasian
	T test	..	..	0.96	..	..
MZ (n=38)	RA (n=19)	66 (8.25)	84	25.44 (6.34)	59* (18.25)	Caucasian
	Control (n=19)	66 (8.25)	84	26.31 (3.79)	60* (18.75)	Caucasian
	T test	..	..	0.96	0.78	..

Table 6.1 RA discordant TwinsUK participant characteristics
The study included 60 participants in total, comprised of identical monozygotic (MZ) and non-identical dizygotic (DZ) twin pairs. *Data unavailable for 3 participants.

6.2.2 Sample processing

Faecal samples were processed as described for the core TwinsUK 16S rRNA dataset in **Chapter 3, Methods 1.** Briefly, on arrival at the laboratory
samples were aliquoted, and frozen at -80°C. Later, DNA was extracted, amplified using PCR primers 805F and 615R, and sequenced on an Illumina MISEQ platform.

6.2.3 Generation of ASVs & Phylogenetic Tree

The 16S rRNA reads were demultiplexed using minimum threshold parameters in QIIME. Files were divided according to per sample forward and reverse reads. Reads were trimmed and filtered according to parameters detailed in **Appendix 3: Chapter6**, and ASVs generated as described in **Chapter 3, Methods 1** for the core TwinsUK 16S dataset. Taxonomic assignment was undertaken using the Silva 1.3.2 reference database (Callahan 2018), and a phylogenetic tree calculated using the Phangorn R package (Schliep 2011), with default parameters.

6.2.4 Statistical Analysis

Analyses were undertaken both using the full cohort, including both MZ and DZ twin pairs, and in MZ twin pairs only. For each type of analysis, the group considered will be specified.

6.2.5 Comparison of alpha diversity in RA affected versus unaffected participants

The same statistical approach for measuring alpha diversity and differences in taxa abundances was used when considering the full cohort (MZ and DZ twin pairs together), and in analysis of MZ twins only. Alpha diversity was calculated using the untrimmed and un-normalised ASV table, as these steps may bias the resultant measures (described in further detail in Chapter 4, Methods 2). In this study sequencing depth was not significantly different between case and control samples (Full cohort, P = 0.1, MZ twins only, P = 0.07) and therefore did not require adjustment. Observed ASVs, Simpson index and Shannon index were calculated using the Phyloseq package in R (McMurdie and Holmes 2013).

In the full cohort, observed ASVs were normally distributed (Shapiro Wilk P = 0.39), whereas Shannon diversity (Shapiro Wilk P = 0.04) and Simpson diversity (Shapiro Wilk P = 2.518e-7) demonstrated a negative skew. As Shannon diversity was almost normally distributed, a transformation was not performed. A square root transformation was applied to Simpson diversity measures, after which the negative skew was abated, but distribution remained non-normal (Shapiro Wilk P = 0.0002). The difference in alpha diversity between RA affected and RA unaffected twins was measured for observed ASVs and Shannon diversity using a paired Student's t-test. For Simpson diversity, the Wilcoxon rank sum test was applied.

When considering MZ twins only, observed ASVs were normally distributed (Shapiro Wilk P = 0.342), whereas Shannon diversity (Shapiro Wilk P = 0.017) and Simpson diversity (Shapiro Wilk P = 2.553e-6) demonstrated a negative skew. A square root transformation was applied to Shannon and Simpson diversity measures, after which the negative skew was abated, but distribution remained non-normal (Shannon, Shapiro Wilk P = 0.006; Simpson, Shapiro Wilk P = 1.806e-6). The difference in alpha diversity between RA affected and RA unaffected twins was measured for observed ASVs using a paired

Student's t-test. For Shannon diversity and Simpson diversity, the Wilcoxon rank sum test was applied.

6.2.6 Taxonomic differential abundance in RA affected versus unaffected participants

Differential abundance of ASVs between RA and control twins was estimated using the DESeq2 R package. Both the full cohort (MZ and DZ twins) and MZ twins only were considered: two sets of analysis for taxa differential abundance were undertaken using similar approaches. The following methods apply to both cohort sets, unless otherwise stated. Filtering of the ASV tables was first applied. ASVs present in less than 5% of samples were removed, in order to focus on taxa which are not overly rare, and in addition as these less abundant taxa are more likely to represent contaminants. Prior to differential abundance estimation, ASVs were transformed using the negative binomial variance stabilising transformation (described in detail in Chapter 4). Regarding covariates, age and sex were inherently accounted for within the twin pairs (and therefore between the groups), as only same sex twin pairs were involved in the study. BMI was not significantly different between case and control groups, and so was not included within the model (Table 6.1). In analysis of the MZ twins, diet was additionally considered. Dietary variance of TwinsUK participants was captured using the healthy eating index (HEI) 2010 (Bowyer et al. 2018). Within MZ twins, the difference in HEI measure between the RA and control twins was not statistically significant (P = 0.78; Table 6.1), and was therefore considered accounted for by the paired twin method and not included within the model.

6.2.7 Abundance of Prevotella spp. between RA discordant twins

As *Prevotella* has been established as a taxon of interest in RA (Scher et al. 2013, Maeda et al. 2016, Larsen 2017, Alpizar-Rodriguez et al. 2019, Iljazovic et al. 2020), I investigated the abundance in our cohort. The following steps were applied both to the full cohort and to the subset cohort containing only MZ twins. ASV counts were converted to relative abundance, where the total for all taxa in a sample sums to 1, and subset to contain only *Prevotella spp.* The relative abundance of all *Prevotella spp.* present in RA compared to control samples and of Prevotellaceae family in RA compared to control samples was calculated and visualised using the Phyloseq R package (McMurdie and Holmes 2013).

6.2.8 Microbiota ecological community relationship and association with rheumatoid arthritis

Statistical approaches have been developed to model the community and niche differentiation of taxa. Construction of balance trees in relation to a phenotype of interest is an elegant approach which allows inference of both community relationship and niche differentiation. Balances - based on degree of proportionality between taxa - originate between pairs of taxa in the dataset, and extrapolate upwards, creating multilevel clades that are inter-related in terms of interaction and niche differentiation. Balances at each level can be investigated in relationship to a phenotype (Morton et al. 2017). Construction of balances sidesteps a key challenge in abundance-based analysis of compositional data. Compositional data varies substantially in the total reads which represent each sample, and consequently the reads for each taxon within the sample. An unknown proportion of the variation is biological, some is artefact secondary to sequencing depth, and additionally taxa with different comparative abundances within samples will respond differently to approaches aimed at addressing compositionality. Further, when considering the change in abundance of two or more taxa between samples, the question arises as to which is

considered the baseline for the relationship. For example, we consider a scenario where duplicate samples were taken containing taxa A and B, representing two time points for the same participant. There are therefore four samples representing two time points, of which 2 are duplicates. At time point two, the proportion of taxon A and taxon B is the same in both of the representative samples, however the counts are different. When comparing these samples, it could be concluded either that taxon A has doubled, or that taxon B has halved. This concept is illustrated in Figure 6.2. The direct comparison of taxon abundances is therefore challenging. The calculation of balances sidesteps this issue. The concept here is that the balance between taxa would be constant, and therefore the question of whether taxon A doubled or taxon B halved is nullified, and the information is captured by the balance. Balances can be calculated to capture the degree of proportionality between taxa, allowing the inference that for taxa which share a balance, the abundance of each is interdependent. Balances may be extrapolated upwards from the initial 2 taxon relationship, to contain multiple sub-balances, representing overall ecological community structure.

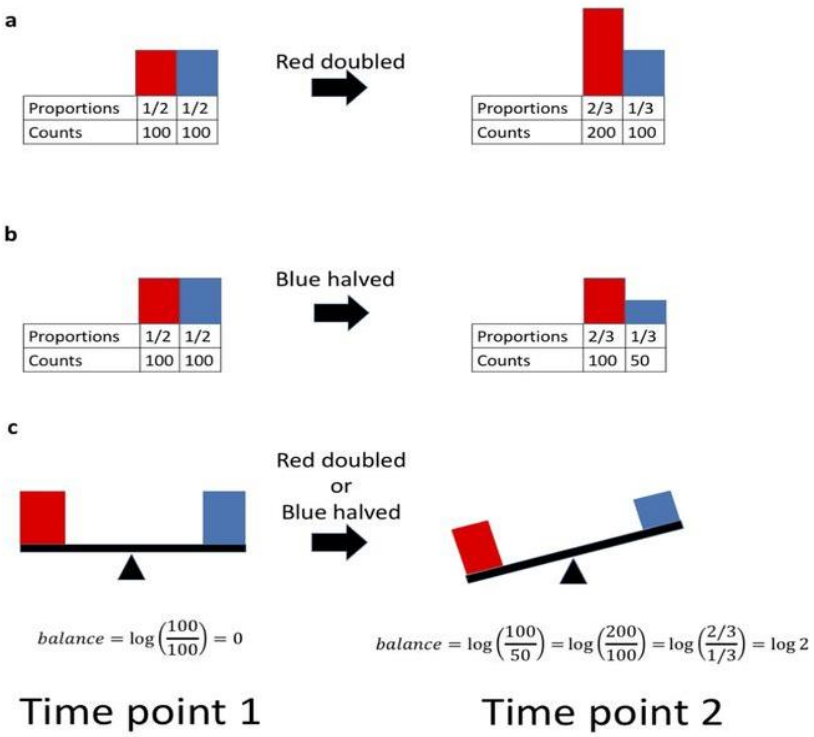


Figure 6.2 Concept of Balances

Hypothetical scenario where 2 samples of 2 taxa in differing proportion may account for different scenarios in the environment (a, b). The balance between the 2 proportions is consistent for both scenarios (c). Adapted from Morton *et al.* (2017).

Balances were generated using the full dataset, including MZ and DZ twins, in order to maximise power. Ward hierarchical clustering was applied to isometric log ratio transformed balances generated using the Gneiss plugin in QIIME2 (Bolyen et al. 2018), after filtering of ASVs present in less than 10% of samples and addition of a pseudo count of 1. The higher filtering threshold used here is in line with the statistical constraints of the method, which employs correlation of taxa with each other. Including rarer taxa would produce spurious results (Morton et al. 2017). Balances are constructed based on the degree of proportionality between features. Within each balance, the proportion of taxon A, located on the left of the balance, is higher compared to taxon B, located on the right. Association of balances with RA diagnosis was assessed using linear models in QIIME 2 (Bolyen et al. 2018, p. 2). If RA (as opposed to genetic potential for RA) creates a changed ecological niche, we may expect to find differences in the balances between established RA patients and their unaffected twins. However given the similarity of the microbiota of established (treated) RA patients to healthy controls demonstrated previously (Zhang et al. 2015), this pattern may not be detected in our cohort of established RA patients.

6.3 Results

RA discordant twin pairs from the TwinsUK cohort attended clinical visits and contributed faecal samples for this study. Several associations were observed between gut (faecal) microbiota variance and RA, both when considering the dataset as a whole - including MZ and DZ twin pairs - and when considering MZ twin pairs only. There were 30 twin pairs in total, comprising 19 MZ pairs and 11 DZ pairs.

6.3.1 Alpha Diversity

There were no statistically significant differences in alpha diversity between RA affected and unaffected participants. This was true when considering the full cohort, MZ twin pairs only or DZ twin pairs only. However, visualisation demonstrated that there is a (non-significant) trend towards lower alpha diversity in RA affected participants compared to their RA unaffected siblings (**Figure 6.3**).

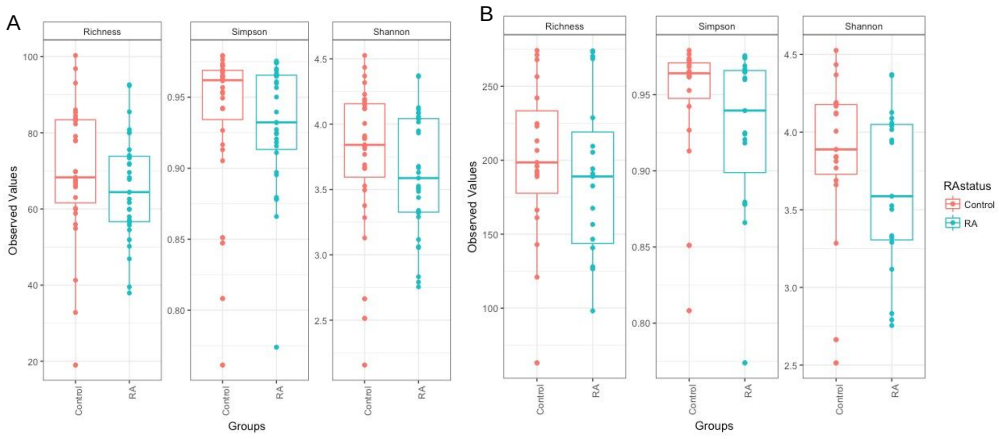


Figure 6.3 Alpha diversity in RA affected participants versus their unaffected twin siblings
In (A) MZ and DZ twins are analysed together (N=60). No significant differences in alpha diversity were demonstrated (Observed ASVs, P= 0.415; Shannon diversity, p = 0.107; Simpson diversity, P = 0.142). In (B) MZ twins only were analysed (N=38). No significant differences in alpha diversity were demonstrated (observed ASVs, P= 0.568; Shannon diversity, P= 0.202; Simpson diversity, P = 0.172). Whilst there is a trend towards lower diversity in RA affected twins compared to unaffected, no measures were statistically different between the groups.

6.3.2 Differential Abundance of ASVs between RA affected and control twins: Monozygotic and dizygotic pairs

Taxonomic differences between the RA affected group compared to the RA unaffected group were demonstrated. There were nine ASVs which were significantly associated with RA when considering the full cohort (N=60 pairs), including both MZ (N=38 pairs) and DZ (N=22 pairs) twins (**Figure 6.4**).

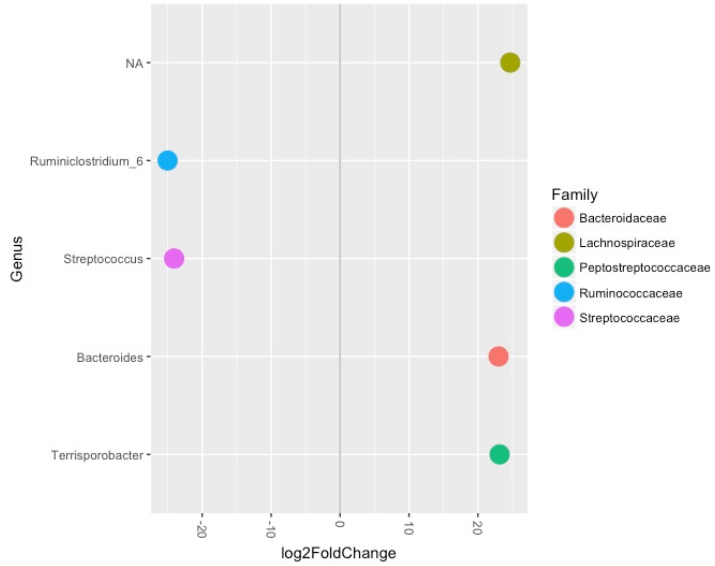


Figure 6.4. Taxonomic association with RA in full cohort: MZ and DZ twins
Differential Abundance of ASVs in RA affected compared to unaffected DZ and MZ twins (n=60). Upregulated in RA samples were *Lachnospiraceae* (genus unassigned) (q = 4.74e-15), *Terrisporobacter* (q = 5.59e-13) and *Bacteroides* (q = 6.9e-13). Downregulated in RA were *Ruminoclostridium_6* (q = 5.06e-15) and *Streptococcus* (q = 5.58e-14).

6.3.3 Differential Abundance of ASVs between RA affected and control twins: Monozygotic twins

There were eight ASVs which were significantly associated with RA when considering the MZ twins only (N=38; Figure 6.5).

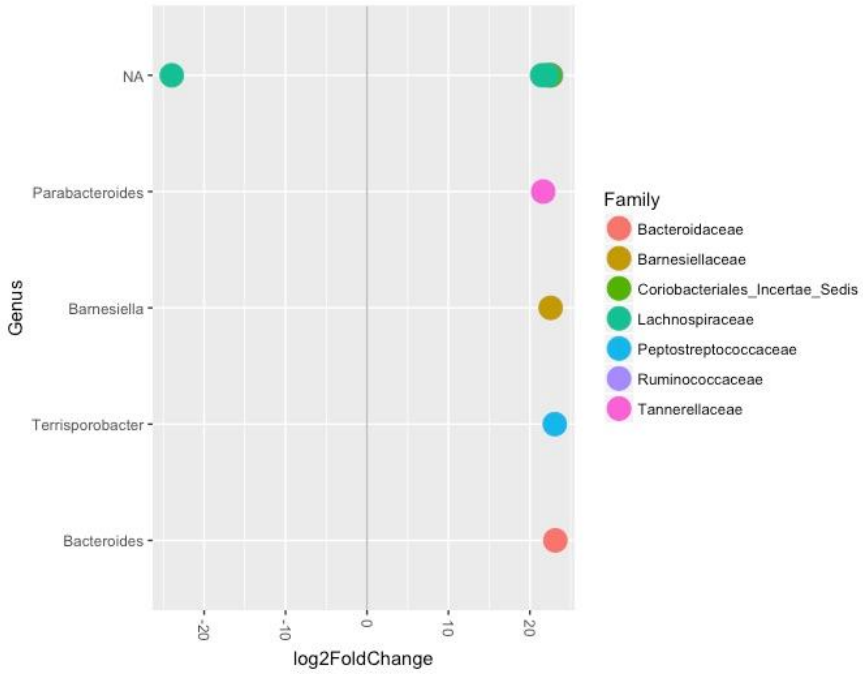


Figure 6.5 Differential Abundance of ASVs in RA discordant MZ twins
Log fold change in abundance in significantly differentially abundant ASVs in RA affected compared to unaffected participants is shown (N=38). ASVs upregulated in RA were *Bacteroides* (q = 2.59e-15), *Terrisporobacter* (3.096e-15), Ruminococcaceae, genus unassigned (q = 1.087e-14), *Barnsiella intestinihominis* (q = 1.24e-14), Coriobacteriales_Incertae_Sedis, genus unassigned (q = 1.366e-14), Lachnospiraceae, genus unassigned (q = 5.41e-12), Parabacteroides distasonis (q = 1.717e-11) and Lachnospiraceae, genus unassigned (q = 2.24e-11).

6.3.4 Abundance of *Prevotella spp*. within RA discordant twins

Motivated by the association of *Prevotella* with RA reported previously, the relative abundance of *Prevotella spp.* in established treated RA versus RA unaffected participants was calculated. No *Prevotella spp*. were significantly different regarding abundance in RA affected compared to unaffected participants. This was true both in all pairs and in MZ twins only. However, variation in abundance of *Prevotella spp.* was suggested. *Prevotella_9*, which is predicted to be *Prevotella copri*, was found to have lower relative abundance in control participants compared to RA affected participants (**Figure 6.6**). When considering all *Prevotella spp*. together at family level, a higher relative abundance of Prevotellaceae was demonstrated in RA unaffected than

RA affected participants (**Figure 6.7**). This suggested difference was not statistically significant.

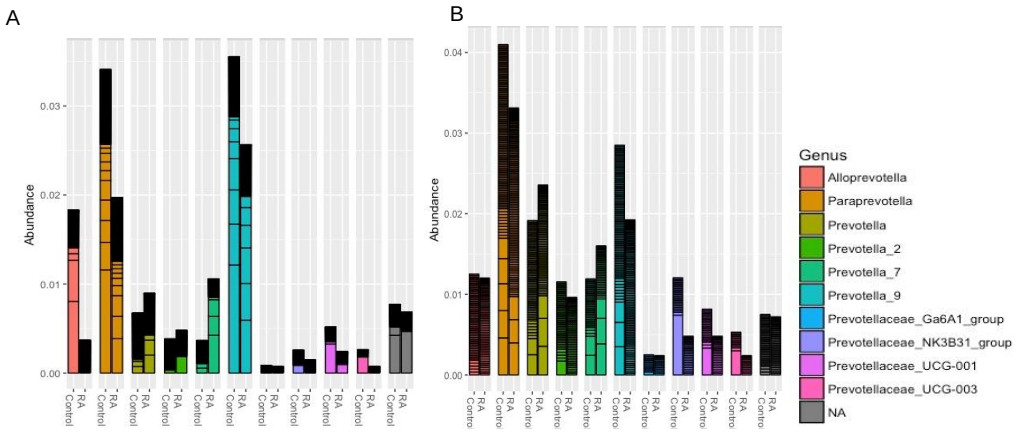


Figure 6.6 Relative abundance of Prevotella spp in RA affected versus control participants
Relative abundance of *Prevotella spp.* in A) MZ and DZ twins (N=60), and B) MZ twins only (N=38). Black lines delineate samples.

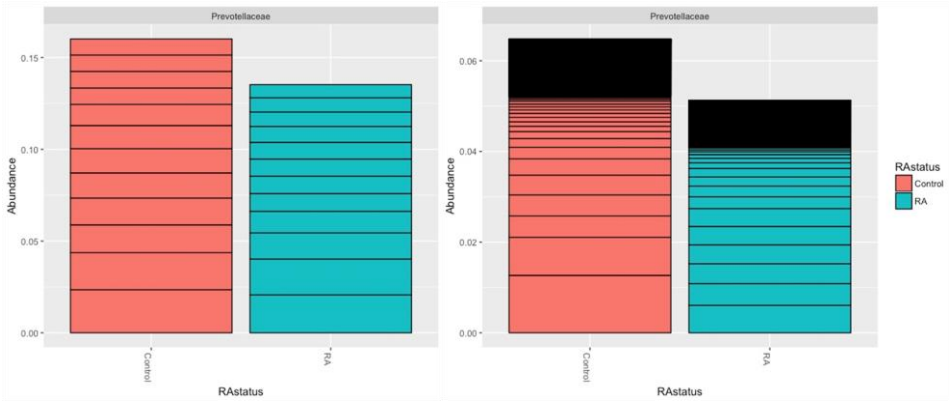


Figure 6.7 Abundance of Prevotellaceae in RA affected versus RA unaffected participants
The relative abundance of Prevotellaceae was higher in control participants than participants with RA. This was true both when considering the full cohort (MZ and DZ twins; B) and when considering MZ twins only (A).

6.3.5 Clustering analysis showing community and niche relationship of core microbiota

Balances represent the ecological community relationship between core taxa . These were calculated in the full cohort, including both MZ and DZ twins . Briefly, there was no difference demonstrated in balances between the group of established RA patients and the group of unaffected twin siblings.

Appendix C shows the taxa which demonstrate compositional interdependence in the dataset. Balances are depicted on the outermost aspect of the tree at the lowest level, containing here 2 taxa per balance. Within each balance, the ratio of the taxon located on the left-hand side of the balance is increased compared to the taxon on the right-hand side of the balance. It is inferred that the abundance of both taxa are inter-dependent, and that both taxa share an ecological niche preference. The above relationship can be extrapolated upwards in the bifurcating tree, in which higher level balances comprising many lower level

balances are depicted. The highest-level balance is Y0, and this contains all of the other balances.

6.4 Discussion

6.4.1 Comparison of the intestinal microbiota of RA discordant twins.

In this chapter, I compared differences in diversity, taxon abundance and community niche differentiation between RA discordant twin pairs. Differences in taxonomic abundance were demonstrated in RA affected verses unaffected MZ twins. Same-sex female RA affected twins were compared as a group to RA unaffected twins with inherent control for age, sex, genotype within the study design. Additional covariates - BMI, diet as measured by HEI and sequencing depth of samples - were not significantly different between case and control groups, and they were not included in the model.

In accordance with other studies, there was a trend towards lower alpha diversity in RA affected versus unaffected participants, however none of the alpha diversity measures demonstrated significant difference between the groups. To date, no prior study has demonstrated a significantly lower alpha diversity in RA affected compared to control participants (Scher et al. 2013, Zhang et al. 2015, Alpizar-Rodriguez et al. 2019), despite large sample sizes, suggesting this is not a power issue. Indeed, many studies demonstrate no difference in alpha diversity between diseased and non-diseased individuals (Reese and Dunn 2018). In the present study, the discordant twin design affords additional power.

The comparison of the intestinal microbiota of monozygotic, same sex RA discordant twin pairs is a strong study design, which circumvents problems of confounding factors, to reveal which In this study, age, sex, genotype, diet, BMI, and early life factors were accounted for within the analysis. This is the first study of the gut microbiota in RA to include consideration of these factors – in prior studies only age, sex and BMI were

considered as covariates. It is therefore interesting to observe that even when the above, including the novel factors of genotype and diet are controlled for, significant taxon differences are observable in the gut microbiota of RA affected participants compared to their RA unaffected twin counterparts. All RA affected twins were assessed against all RA unaffected twins, as opposed to paired testing. Taxa which were associated with RA in monozygotic RA discordant twin pairs were *Bacteroides*, *Terrisporobacter*, *Barnsiella*, *Terrisporobacter*, *Parabacteroides*, *Lachnospiraceae spp.* and *Cioriobacterales Insertae-Sedis spp.*

Intriguingly, despite having the most carefully considered study design affording control of additional important factors, this is the first study to report significant differences in between the intestinal microbiota of established (treated) RA patients and healthy control participants. This may be due in part to prior studies having focussed of treatment naïve RA, and the broad shift in intestinal microbiota towards a more "normal" composition in those undergoing treatment with DMARDS (Scher et al. 2013, Zhang et al. 2015). Thus, prior work has focussed on similarity as opposed to differences in established RA patients. The current study had comparable sample size, having 30 established (treated RA participants), where the prior studies included 28 and 21 established RA patients. Only the study by Scher *et al.* included a differential abundance analysis (using LEFSE) between established RA participants and controls, and they did not report any significant difference.

There are some key similarities, but also notable differences, between the findings of our study and those previously reported. In the metagenomic study by Zhang *et al.*, the gut microbiota of treatment naïve Chinese RA patients was compared to that of control participants. Bacterial families enriched in the gut microbiota of RA patients were, in order of abundance, Bacteroidaceae, Lactobacillaceae, Clostridaceae, Coriobacteraceae and Erysipelotrichaceae. Taxa represented by metagemonics linkage groups (MLGs) which were enriched in RA patients or controls are presented in Table 6.2.

RA Group: New onset (un-treated) / Established (treated)	Higher abundance in RA	Higher abundance in Controls	Ref.
New-onset (n = 44) Healthy Controls (n = 28)	*Prevotella copri*	Not reported.	(Scher et al. 2013)
New-onset RA (n = 94) Healthy Controls (n = 97)	*Bacteroides*, *Bacteroides stercoris ATCC 43183*, *Parabacteroides, Bacteroides vulgatus ATCC 8482, Bifidobacterium dentium JCVIHMP022, Burkholderiales bacterium 1_1_47, Citrobacter freundii 4_7_47CFAA, Citrobacter sp. 30_2, Clostridiales sp. SS3/4, Clostridium asparagiforme DSM 15981, Clostridium glycolicum ATCC 14880, Clostridium leptum DSM 753, Clostridium scindens ATCC 35704, Clostridium sp M62/1, Coprobacillus sp. 29_1, Eggerthella Lenta, Enterobacter cloacae, Enterococcus faecium.* Association of *Prevotella copri* with RA severity in first year of clinical onset.	*Bacteroides plebeius M12, DSM 17135, Bifidobacterium bifidum, Capnocytophaga sp. oral taxon 329 str. F0087, Clostridales spp., Clostridium saccharolyticum WM1, DSM 2544, Collinsella aerofaciens ATCC 25986, Coprococcus catus GD/7, Coprococcus sp. ART55/1, Dialister invisus DSM 15470, Dorea formicigenerans ATCC 27755, Enterobacter cloacae cloacae NCTC 9394* and *Haemophilus parainfluenzae T3T1.*	(Zhang et al. 2015)
Established (n = 26) Healthy Controls (n = 28)	No significant differences reported. Higher abundance of *P.copri* in healthy controls versus established treated arthritis (ns). Abundance of	No significant differences reported.	(Scher et al. 2013)

	Bacteroides versus Prevotella described as determining feature between new-onset RA and both treated RA and healthy controls. Abundance of Bacteroides highest in established RA, but not significant when compared to either other group.		
Established RA (n = 21) Healthy Controls (n = 97)	No significant differences reported.	No significant differences reported.	(Zhang et al. 2015)

Table 6.2 Summary of findings of prior studies of the gut microbiota of established (treated) RA patients

Scher *et al.* (2013) undertook a study using both 16S and metagenomes to assess the abundance of the gut microbiota in new onset RA, chronic treated RA, psoriatic arthritis, and healthy controls. They showed that there was a higher abundance of *Prevotella spp.* (OTUs assigned to genus level to taxa which belong to Prevotellaceae) in new-onset (untreated RA) compared to healthy controls. Conversely, when comparing chronic treated RA with healthy control participants, they showed a higher abundance of *Prevotella spp.* in healthy controls compared to chronic treated RA participants. By undertaking a metagenomic analysis in a subset of participants, they were able to identify that the most abundant *Prevotella* assigned OTU was *Prevotella copri.* They were therefore able to extrapolate this taxon annotation using metagenomics in a cohort subset to the 16 data representing all study participants. They identified that the number of samples containing *Prevotella copri* was 6/28 in healthy controls and 3/26 in chronic treated RA. These data are presented in Table 3, Scher *et al.* (2013). In accordance with this study (shown in **Figure 6.7**) the difference in abundance of neither *Prevotella spp., n*or *Prevotella copri* specifically reached statistical significance. However, it is interesting to note the finding of lower abundance of *Prevotella spp.* is replicated across their study and ours. Potentially, treatment

with immunomodulatory medication which suppresses the inflammatory response may lead to a lower abundance of *Prevotella* in chronic treated RA versus healthy controls. It would be of interest to study people who have not taken DMARDs for whatever reason.

In the study by Scher *et al.,* abundance of *Bacteroides* was low in new onset RA patients but dominated samples from healthy controls and from chronic treated RA participants. In accordance with the results of this study, the abundance of *Bacteroides* was increased in chronic treated RA participants compared to healthy control subjects (Supplementary Figure 1D, Scher *et al.* 2013), but did not reach a statistically significant difference (Scher et al. 2013). They used primers targeting the V1-V4 regions, and sequencing was via the Roche 454 platform.

The higher abundance of *Bacteroides* in established, treated RA patients may have relevance to RA pathogenesis. In the context of commensal intestinal microbiota, *Bacteroides* have been demonstrated to have beneficial immunomodulatory function. *Bacteroiodes fragilis,* a common commensal and the dominant *Bacteroides* species, has demonstrated anti-inflammatory immunomodulatory function and has previously been hypothesised to be implicated in RA (Wu et al. 2016). Polysaccharide A derived from *Bacteroides fragilis* has been demonstrated in human as well as murine studies
In support of above hypothesis, a lower level of *Bacteroides* is a feature of inflammatory bowel disease (Zhou and Zhi 2016).

Overall, there are insufficient studies to establish whether uniform observations of taxon associations are observable during the course of RA. Adding complexity to these observations is that prior studies have implemented different methods and included participants of different ethnicities. However there is agreement that specific compositional differences in the intestinal microbiota between new-onset RA and established RA cases exist, with the latter having a composition more similar to healthy control participants than other (untreated) RA patients. The results of this study highlight that there may be differences, previously overlooked, in the intestinal microbiota of treated RA patients,

which are associated with the disease itself, or it's treatment. The latter is more likely, but longitudinal studies would be required to confirm this. At this stage, further studies are required to confirm taxon associations with RA.

In the present study community composition was not observed to differ in RA. Microbiota balances, composed of core ASVs, were not significantly different between RA and control twins (Appendix D). However it is not possible to conclude from our results that there is no association between ecological community relationship and RA. We included only RA discordant twin pairs in the study, who are difficult to source; this has been reflected in the sample size and as such this could be power issue. Additionally, due to the constraints of the method, a higher filtering threshold was applied for the community analysis. While some taxa identified in the discordant twin analysis were present, and indeed, grouped together in balances representing microbial communities, some of the taxa which were associated with RA in the discordant twin analysis are not included in the community analysis. There may be differences in community structure between RA and control participants which were not demonstrated in this study.

An additional observation from the microbial community analysis is in relation to *Bacteroides* and *Prevotella*, which have been suggested in prior studies to be antagonistic with each other (Scher et al. 2013, Gorvitovskaia et al. 2016). These are dominant taxa, which tend to comprise a substantial portion of the gut microbiota, and as such have been posited as representative taxa of opposing enterotypes (Roager et al. 2014). Antagonism is a category of community relationship and therefore it would be expected that taxa which are antagonistic would group together in the same lower level balance, suggesting either a direct relationship or shared environmental niche between the two taxa. *Bacteroides* and *Prevotella* did not group together within the same balance until higher up in the bifurcating tree, at balance Y6 (Appendix C). Therefore, the antagonistic nature of *Bacteroides* and *Prevotella* is only tentatively supported. This is of relevance to RA, as these taxa are associated with early onset treatment naïve, and established treated RA. *Prevotella* has been shown to dominate the gut microbiota of treatment naïve RA patients (Zhang et al. 2015),

whereas *Bacteroides* dominates the gut microbiota of established treated RA patients as shown in the present study (Figure 6.4 and Figure 6.5) and in prior studies (Scher et al. 2013, Zhang et al. 2015).

6.4.2 Limitations

The first limitation of this study is that the RA affected participants were undergoing treatment with DMARDs. Medication is known to affect the composition of the microbiota. In RA, the pattern is different to that which is usually found. As discussed, prior studies have indicated that the microbiota of RA patients after they begin treatment with DMARDs is actually largely similar to healthy control participants, and more so than when compared to other RA patients who are untreated.

A second limitation of this work is relatively small sample size of 60. It is challenging to source RA discordant twin pairs for such a study and we were fortunate in having access to TwinsUK with its longitudinal data collection over 20+ years (Verdi et al. 2019). Even with this constraint, our sample of established RA patients is larger than both previous studies, having sample sizes of 28 (Scher et al. 2013) and 21 (Zhang et al. 2015). The more stringent filtering of lower abundance taxa required is one limitation of the balances and other clustering methods. Low abundance taxa may be of interest in relation to disease phenotypes but would not be included in these analyses as it would not be statistically appropriate.

The general limitations of 16S sequencing data are given in Chapter 8, Discussion. Briefly, it is not possible accurately to infer the microbial genes present and it is usually not possible to assign taxonomy below genus level. However, in some instances accurate species assignment has been possible secondary to the use of ASVs as opposed to OTUs.

6.4.3 Future work

Regarding future work there are a number of possible approaches that could be implemented. Firstly, a paired analysis of taxonomic differential abundance is possible using the EdgeR package in R. This would allow for paired comparison of taxon abundances between RA discordant twin pairs. Potentially there is some justification for using the DESeq2 approach given that multiple taxa perform the same function it is more useful to detect taxa which are differentially abundant between RA affected and unaffected groups as opposed to a paired analysis - paired differential abundance. A paired analysis of genes present and their associated functions would be informative but is not possible using 16S data.

It would be helpful to explore if the RA affected twins are on similar or different medication regimes for the treatment of their RA. In prior studies, as discussed above, DMARD treatment for RA is associated with fewer microbiome differences in RA patients compared to healthy controls. However, the study is likely to be underpowered to demonstrate association of the microbiota with RA treatment; the study is being expanded to include a set of newly diagnosed RA patients (sufficient twin pairs would be impossible). There is also suggestion that the microbiome is associated with differing response to DMARDs. This requires extensive further investigation and is the subject current projects.

It will be important to look at association of RA clinical parameters with the gut microbiome in the RA affected twins. For example, if one obtained measurements of erythrocyte sedimentation rate (ESR) or C reactive protein (CRP), it would be possible to calculate DAS28 and assess for correlation with microbiota. Potentially taxa of interest may not be differentially abundant between RA affected and unaffected twins but may correlate with disease activity. It would also be helpful use metagenomics methods, which are being planned for the extension of this study.

The balances could be calculated in MZ twin pairs only. Although this analysis would have lower power, it would hold the advantage of controlling for host genetic factors. However, as the results for this work were negative, furthermore stringent analysis may not be informative. Alternatively, it would be interesting to investigate in the RA affected twins association of balances with DAS28.

6.4.4 Conclusions

Prior studies have focussed on the gut microbiota of new onset treatment naïve RA cases. However the present study demonstrated significant differences in established RA cases compared to controls. There were detectable taxon differences in the gut microbiota of RA affected twins, when compared to non-affected twins. Thus, taxon differences were detectable even when host genotype and early life factors shared between twin pairs, are controlled for. In accordance with prior studies, there was no association between *Prevotella* and established RA. *Bacteroides,* having been shown previously to dominate the gut microbiota of established RA patients, and to have anti-inflammatory immunomodulatory influence, demonstrated the most marked higher abundance in RA participants. However the species of *Bacteroides* was not identified in the present study, and results of this study were therefore inconclusive with regards to implication of the microbiome in RA pathogenesis.

Chapter 7

Association of Host Genetic Factors with the Oral Microbiota in Rheumatoid Arthritis

In this chapter, I consider the link between oral microbiota, periodontal disease and RA. I investigate the association of the saliva microbiota with genetic risk of RA; as part of this, I demonstrate heritability of taxa of the saliva microbiota using classical twin modelling. In order to both contextualise findings of the present study, and to provide a wider contribution to the research area, I characterise the saliva microbiota including relation to a range of potentially influential biological and technical factors.

Collaborator Attributions

Daniel Sprockett, collaborator of Claire Steves and myself, performed pre-processing and sequencing of saliva samples and generated ASVs for this dataset at the University of Stanford. Ruth Bowyer generated the frailty and healthy eating indices.

7.1 Introduction

7.1.1 Background

The oral microbiota - colonising the saliva, teeth, and buccal mucosal surfaces - has been of particular interest in RA, alongside the gut and pulmonary microbiota (Scher et al. 2012, Zhang et al. 2015). There is an intriguing suggested relationship between the oral

microbiota and the development of RA (Bingham and Moni 2013, Eriksson et al. 2019). First, periodontal disease - associated with characteristic differences in the oral microbiota - is a major risk factor of RA. Second, bacteria which associate with periodontal disease may aggravate RA pathology in an aetiologically specific manner. Third, periodontal disease and RA, both being inflammatory mediated conditions, may share an overlap in genetic pre-disposition. Fourth, periodontal disease and RA share influential environmental risk factors such as smoking. Finally, both conditions share pathophysiologic similarity of systemic and local inflammation and bone erosion, affecting the joints in RA and the oral cavity in periodontal disease. In support of this, in patients with both periodontal disease and RA, severity of bone erosion at both sites, teeth and joints, has been shown to be correlated (Marotte et al. 2006).

7.1.2 Association of the oral microbiota with RA

The oral microbiota exert systemic influence and may have a role in RA. For example, within the oral microbiota a candidate keystone taxonomic association with RA is *Poyphyromonas gingivalis* (*P.gingivalis*) (Ceccarelli et al. 2018). *P.gingivalis* has been suggested to contribute to RA pathology - *P.gingivalis* may upregulate RA specific autoantibodies. *P.gingivalis* possesses peptidyl arginine deiminase (PAD), which has the uncommon ability to citrullinate proteins, providing a precursor step for increased production of anti-citrullinated protein antibodies (ACPA) – a cardinal feature of RA. To date no other bacteria have been demonstrated to possess PAD, and potentially this enzyme is unique to *P.gingivalis* (Vermilyea et al. 2019). A hypothesis has been put forward that ACPA raised in response to citrullinated proteins within the oral cavity may translocate to the joints through epitope spreading, and contribute directly to synovium inflammation, osteoclastogenesis and bone erosion. In this scenario, local inflammation in the oral cavity associated with periodontitis may compromise the mucosal barrier, allowing ACPA to be transported from the oral cavity to other sites via the bloodstream (Kurowska et al. 2017).

There is some evidence to suggest this may be feasible. In a murine collagen-induced model of arthritis, oral administration of *P.gingivalis* but not other bacteria resulted in joint inflammation (Jung et al. 2017). An alternative and perhaps more plausible or likely explanation is that there may be an influence of ACPA raised in response to *P.gingivalis* citrullinated proteins on other immune components, which in turn contribute to RA pathology (Holers et al. 2018). An influence of *P.gingivalis* via the gut microbiota has also been suggested (Sato et al. 2017). Overall, an influence of *P.gingivalis* on RA pathophysiology, secondary to upregulation in periodontal disease and perhaps unique ability of protein citrullination, is a compelling hypothesis, however it may not hold up under scrutiny of future studies.

A concurrent mechanism which links *P.gingivalis* with RA, is production virulence factors, known as gingipains (also described in Chapter 1, Section 1.3.6) in relationship to Alzheimer's disease). Gingipains are cysteine proteases which are transported in outer membrane vesicles (OMVs), conferring the possibility of inter-site transmission without the need for bacterial translocation (Dominy et al. 2019). The relevance of gingipains is two-fold. Firstly, they are a co-factor in citrullination via PAD (Möller et al. 2020). Secondly, antibodies to arginine gingipain (RgpB) have been demonstrated to associate with both periodontal disease and RA (Kharlamova et al. 2016). Anti-RgpB IgG was significantly upregulated in periodontal disease patients compared to controls, and in RA patients compared to controls. Regarding RA, the association was stronger in participants who were seropositive for ACPA, compared to those who were seronegative. Surprisingly, the study showed that the effect size of the association between RgpB IgG and RA (OR=2.96; 95% CI: 2.00–4.37) was stronger than for the association between smoking and RA (OR=1.37; 95% CI: 1.07–1.74). In addition, they demonstrated an interaction between RgpB antibodies with both shared epitope positivity and smoking. This supports the suggestion above that any interaction between RA pathology and *P.gingivalis* extends beyond PAD enzymes. Another mechanism may be translocation of proteins citrullinated by *P.gingivalis* in the oral cavity, from the oral cavity to the bloodstream, and subsequent generation of circulating auto-antibodies.

In support of this link, a recent study of non-RA participants with positive ACPA (thus likely on the pathway of RA development and classified as at risk) demonstrated increased prevalence of plaque *P.gingivalis* at healthy non-inflamed gum sites compared to both healthy controls (effect size, 3.00; 95% CI, 1.71-4.29) and to participants with early RA (effect size, 2.14; 95% CI, 0.77-3.52) (Mankia et al. 2019). Whilst the study reports that healthy oral sites were sampled, the pattern was similar to periodontal disease prevalence in each group: 73% in at risk participants, 38% in healthy controls and 54% in early RA participants.

Evidence for an association of *P.gingivalis* in the oral microbiota with RA or periodontal disease is not easy to disentangle. There are conflicting findings relating to the abundance of *P.gingivalis* in RA patients compared to controls in the two major prior studies published to date on the oral microbiome in RA (Scher et al. 2012, Zhang et al. 2015). In the study by Zhang *et al.,* no association of saliva or plaque *P.gingivalis* with RA was seen, but conversely, a significant enrichment in *P.gingivalis* from both sites in control participants was demonstrated. In the study by Scher *et al.,* there was no association demonstrated for *P.gingivalis* with RA or controls. Clearly, more studies are needed to investigate whether *P.gingivalis* is implicated in RA. Potentially, this taxon could be pathologically relevant in RA development, but not found in higher abundance in established RA patients compared to controls; these two assertions are not mutually exclusive. Studies linking abundance of *P.gingivalis* with RA clinical parameters suggest that although *P.gingivalis* may not be upregulated in RA, it could nevertheless aggravate disease physiology in genetically susceptible individuals (Kharlamova et al. 2016, Sherina et al. 2019).

Whilst there has been a focus on *P.gingivalis* as a link between the oral microbiota, periodontitis and RA, it is more plausible that other taxa are implicated. Indeed, there are numerous other taxon associations between the oral microbiota and both periodontal disease and RA. Regarding periodontal disease, in addition to the traditional 'red complex' of *P. gingivalis, Treponema dentolica,* and *Tannerela forsythia,* taxa which potentially possess a

direct aetiological link with RA also include *Aggregatibacter actinomycetemcomitans (Aa)* and *Prevotella intermedia* (Mankia et al. 2019). *Aa* was demonstrated to induce the release of pore forming toxins, leading to neutrophil release of citrullinating enzymes. The same pore forming toxin induced a change in neutrophil morphology, leading to the release of citrullinated peptides (Konig et al. 2016). An immune response to *Prevotella intermedia* was associated with antibodies to specific citrullinated peptides in gingival crevicular fluid of participants with RA and periodontal disease (Schwenzer et al. 2017).

Regarding taxon associations of the oral microbiota with RA, inverse associations include *Veillonella* and *Haemophilus*, *Klebsiella pneumoniae*, *Bifidobacterium bifidum*, *Sutterella wadsworthensis* and *Megamonas hypermegale*. Positive associations include *Clostridium asparagiforme*, *Gordonibacter pamelaeae*, *Eggerthella lenta* and Lachnospiraceae bacterium, *Bifidobacterium dentium*, *Lactobacillus sp.* and *Ruminococcus lactaris* (Zhang et al. 2015). Mechanisms by which these other taxa may contribute to RA pathology have not been widely discussed. However, a general hypothesis is that oral bacteria can influence the systemic immune state (Belkaid and Hand 2014). For example, transport of persistent bacterial components in DC and other phagocytosing cells and shaping of T cell memory at mucosal sites. Such events may be exacerbated by disruption of the mucosal membrane, which occurs more frequently in periodontal disease (Möller et al. 2020).

Until recently the oral microbiota were less studied in comparison to the gut microbiota in relation to RA, although both the oral and gut microbiota may hold important roles in disease pathogenesis. In addition, the oral microbiota may provide a more accessible clinical target compared to the gut microbiota, as it is based at the end of the gut rather than the middle. A prior study demonstrated amelioration severity of RA via periodontal therapy (Ortiz et al. 2009).

7.1.3 Influence of host genetic factors on the oral microbiota

Host genotype is demonstrated to influence oral microbiota composition and genetic risk factors in RA may act partially via the oral microbiota. Prior studies have demonstrated that the oral microbiome is highly heritable. Whilst the heritability of the adult oral microbiome is yet to be characterised, a study in children of sub-gingival plaque showed that some OTUs were as much as 65% heritable (Gomez et al. 2017). The most heritable OTU was reported to be *Prevotella pallens*, followed by *Veillonella* (60% heritable) and *Corynebacterium durum* (54% heritable). The difference in these reported heritable taxa of the oral microbiota compared to the heritable taxa of the intestinal microbiota (Chapter 3) is expected, as the composition of the oral and gut microbiota are influenced by different factors. Influence of host genetic factors of the gut microbiota (investigated in Chapter 5) is likely to be driven primarily by the immune milieu. This is also likely to hold true for the oral microbiota - immune components within the oral cavity are likely to influence the composition of the oral microbiota. However there are additional factors which may influence the composition of the oral microbiota, which would be secondary at least in part to genetic factors: enzymes such amylase and lysozyme; proline-rich proteins; saliva flow rate which affects pH and ion concentrations; small molecules present in saliva such as histatins and cystatins; mucins and taste receptors (Davenport 2017). Further studies of the heritability of the adult microbiome are required.

The influence of host genetic factors on the oral microbiome in RA has not been investigated previously. In this chapter I investigate whether published RA risk loci are associated with oral microbiota differences in unaffected individuals, and whether these are the same as seen in RA patients. In order to isolate genetic association with the microbiota from RA disease physiology and treatment, which may have their own effects on the oral microbiome, exclusion criteria of RA diagnosis or twin sibling with RA were applied. Genetic risk of participants of RA was determined using polygenic risk scoring, and association assessed with the oral microbiota.

In this chapter I show that elements of the salivary microbiota are heritable. I go on to hypothesise that host genetic factors influencing the salivary microbiome may be common with RA.

7.2 Methods

7.2.1 Participants

Participants of this study are members of the TwinsUK cohort (Verdi et al. 2019) who had been previously genotyped. The majority of participants were Caucasian females. The age of the participants ranged from 38 to 80 (median age 68). General health of the participants was captured using a frailty index generated from self-report diagnoses, following the Rockwood frailty index method (Searle et al. 2008). Briefly, this is a measure of health deficit computed by dividing the number of age-associated health deficits by the total number of domains. Participant demographics are summarized in Table 7.1.

	Age median (IQR)	Sex (% F)	BMI median (IQR)	HEI median (IQR)	Frailty median (IQR)	Zygosity	Ethnicity	No. days samples stored median (IQR)	Fasted Status (% +ve)
TwinsUK participants (N=407)	68 (12)	95	26 (6)	55 (11)	0.18* (0.13)	MZ: 166 DZ: 241	Caucasian: 402 Mixed: 2*	325 (610)	93

Table 7.1 Participant Characteristics
* Data unavailable for some participants

7.2.2 Polygenic Risk Score for RA

The polygenic risk score for RA was generated as described in Chapter 5. The same methods for polygenic risk scoring were used for the study of the gut microbiota, and the present study. Briefly, genome-wide significant single nucleotide polymorphisms (SNPs) published as associated with RA were identified using the NCBI generated database of GWAS summary statistics. An inclusion criterion of European ancestry ensured studies were representative of TwinsUK demographics. Risk allele dosage of the identified SNPs were extracted within TwinsUK data using Plink 1.9. Pruning was applied to account for linkage disequilibrium. The risk allele frequency was then multiplied by the SNP-association beta, to produce a weighted PRS. The PRS was tested for its predictive value for RA in 6,776 participants from UK Biobank – 2,686 RA cases, 4,090 controls. Diagnosis of RA in the UK Biobank participants was determined using hospital episode statistics (HES) data supplied by NHS Digital, and logistic regression of RA cases and unselected controls with no evidence of HES diagnosis of RA, against PRS - adjusting for age, sex and smoking history - was applied.

7.2.3 Saliva sample collection

Saliva samples were collected from participants during routine volunteer visits to the Clinical Research Facility associated with the Department of Twin Research at King's College London. Participants were requested to arrive for their volunteer visit to the Clinical Research Facility having fasted for six hours. Prior to collection of saliva samples, participants abstained from food, drinking beverages, smoking, and chewing gum for at least one hour, and the majority were fasted for more than 6 hours. Participants were instructed to spit into a 30 ml sterile Falcon tube for ten minutes, and to try and produce as much saliva as possible in this time. To aid saliva production, participants were instructed to think about their favourite foods whilst providing the sample. Completed samples were

immediately placed in a refrigerator before being transported to the laboratory in an insulated cooling bag. On arrival in the laboratory, samples were aliquoted into Eppendorf tubes and frozen at -80°C for storage purposes. Frozen saliva samples were shipped on dry ice to Stanford University for DNA extraction and 16S rRNA gene sequencing.

7.2.4 Saliva sample processing and sequencing

For the following steps, undertaken by Daniel Sprockett, University of Stanford in collaboration with Claire Steves, saliva samples were randomly distributed, as opposed to being organised according to participant ID, as this may introduce bias due to twin pairs' samples being placed adjacent to one another. Both DNA extraction and PCR blanks were included. DNA extraction was performed using the DNeasy PowerSoil HTP 96 DNA extraction kit (Qiagen) according to the manufacturer's instructions. This includes a mechanical cell lysis step, of bead beating for 20 minutes.

The V4 region of the 16S rRNA gene was amplified using PCR, in triplicate, using primers 515F and 806R which include error-correcting barcodes and illumine adaptors. Prior to sequencing, samples were pooled in equimolar ratios. Sequencing was undertaken on an Illumina HiSeq 25000 platform, which generated a total of 167.8 million reads. Following sequencing, data were returned to the Department of Twin Research.

7.2.5 Microbiota profiling: Generation of ASVs

Reads were denoised using DADA2 to generate ASVs (Callahan et al. 2016). A similar approach was taken as for the TwinsUK intestinal ASVs. Chimeras were removed using the consensus method (described in Chapter 3). Taxonomy of ASVs was assigned using the SILVA database, version 1.3.2 (Callahan 2018). ASVs assigned as "mitochondria" or "chloroplast", or which were unassigned at Kingdom level were removed from the dataset.

DNA extraction and PCR blanks were utilised to detect contaminant ASVs, which originate from laboratory equipment, laboratory reagents or personnel, as opposed to the biological samples. DNA extraction and PCR blanks were used as input for the Decontam R package version 1.2.1 (Davis et al. 2018), which identified 46 contaminant ASVs. The ASVs identified as contaminants were removed from the dataset. A phylogenetic tree was generated for the ASVs by matching the ASVs into the SILVA nr v132 phylogenetic tree backbone using the fragment-insertion function (version 2018.6.17) in QIIME2.

7.3 Statistical analysis

Taxon composition of the saliva microbiota was visualised using the Phyloseq R package, after conversion of taxon counts to per sample relative abundance. Alpha diversity was calculated from the untrimmed and un-normalised ASV table (McMurdie and Holmes 2014) using three measures – Shannon Index, Simpson Index and Observed ASVs. Bray-Curtis distance was calculated after applying the negative binomial variance stabilising transformation, to account for distribution and sequencing depth (described in Chapter 4).

Heritability estimates for the saliva microbiota followed the approach detailed in Chapter 3. Heritability was calculated for ASVs present in more than 15% of samples from 344 twin pairs (192 MZ pairs, 152 DZ pairs). Residuals derived from linear mixed effects models of relative abundance of ASVs, covarying for age and sex as fixed effects and sample storage time and sequencing depth as random effects, were box-cox transformed and used as input for heritability estimation.

Linear mixed-effects models were used to assess association between PRS for RA and alpha diversity. Models were run with alpha diversity as a response variable to the PRS. Models were run using the lme4 package in R, (Bates et al. 2015) with fixed effect covariates age, BMI, diet (HEI), frailty, fasting status, storage time of sample and sequencing

depth, and family structure as random effects. Seven samples with data thought to be missing at random for frailty were excluded from the analysis. Standardised coefficients are reported.

Association of RA PRS with beta diversity was calculated using permutational analysis of variance (PERMANOVA) applied via the Vegan R package (Oksanen et al. 2019), covarying for age, sex, BMI, HEI, frailty, fasting status and time in freezer. The PERMANOVA was repeated using each variable as the primary explanatory variable, and the remainder as covariates. Data thought to be missing at random for frailty, meant models were run twice, on the full dataset not including frailty, and in the subsample for which frailty score was available, with adjustment for frailty.

Difference in abundance of ASVs present in more than 40% of samples against the PRS was assessed using DESeq2 (Love et al. 2014b) covarying for age, sex, BMI and family structure. The higher threshold was chosen as the sample size for the present study was small, in order to include taxa which are more likely to be found in the wider population. Adjustment for multiple testing was applied to all models, using false discovery rate (FDR).

7.4 Results

7.4.1 Taxon characterisation of the saliva microbiota

Within 407 participants the saliva microbiota comprised 2,602 ASVs, comprising 15 phyla (**Figure 7.1**). Of these,155 ASVs were present in more than 40% of samples, comprising of ten phyla (**Figure 7.2**). There were 64 ASVs present in more than 80% of samples, comprising 7 phyla (**Figure 7.3**).

Across both filtered and unfiltered samples, the most dominant phyla in ascending order were Proteobacteria, Firmicutes and Bacteroidetes. Within the total dataset, the top 20%

most abundant taxa were *Haemophilus, Streptococcus, Veillonella, Prevotella_7, Lactobacillus, Neisseria, Rothia and Fusobacterium* (Figure 7.5).

Figure 7.1 Relative abundance of ASVs within the saliva microbiota at the phylum level (n=407).

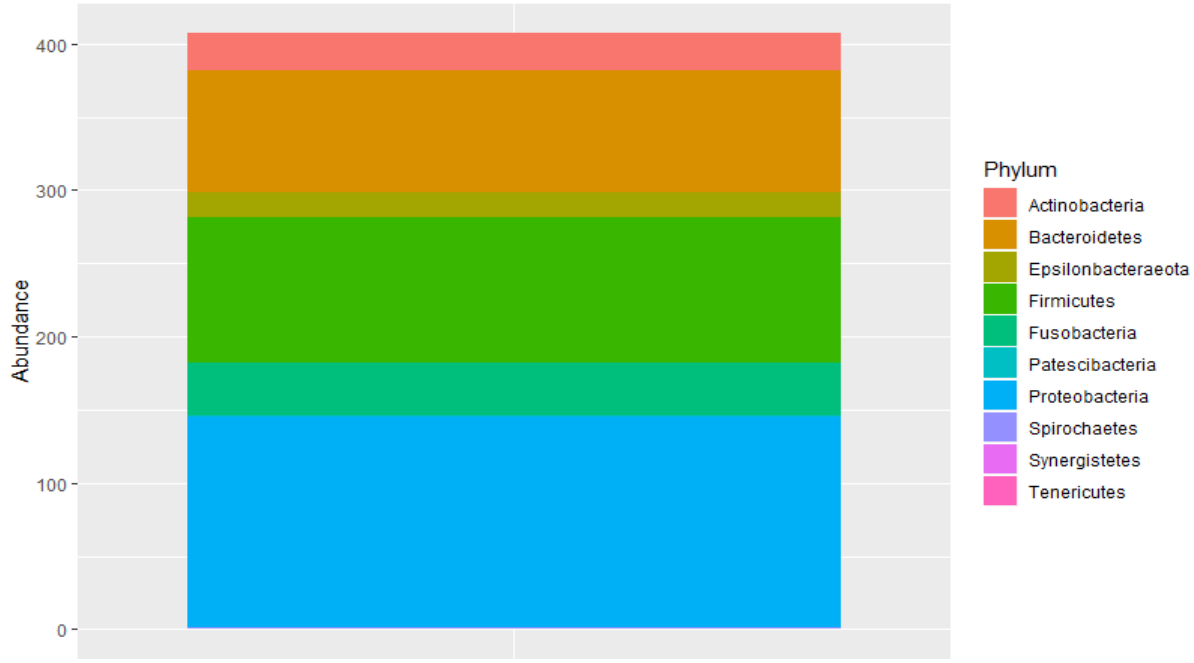


Figure 7.2 Relative abundance of ASVs present in more than 40% of saliva microbiota samples, at the phylum level
(n=407)

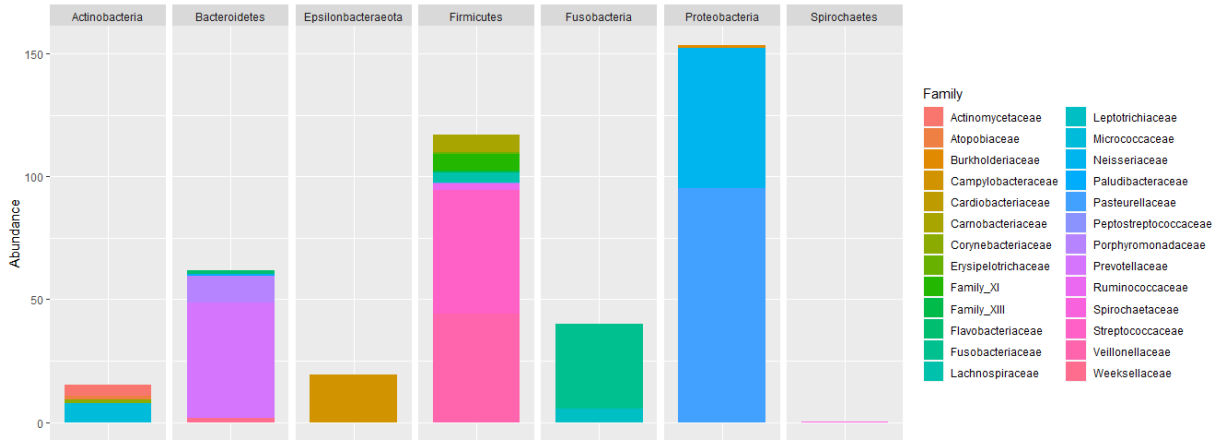


Figure 7.3 Relative abundance of ASVs present in more than 80% of saliva microbiota samples, at the phylum level in TwinsUK
(n=407)

The top 20 most abundant taxa across all samples comprised 63 ASVs assigned to seven phyla: Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Fusobacteria, Epsilonbacteraeota, and Spirochaetes (**Figure 7.4**).

At genus assignment level, the most prevalent taxa in the cohort, defined as presence in more than 80% of samples, were *Streptococcus, Haemophilus, Veillonella, Prevotella_7, Prevotella_6, Ruminococcaceae-UCG-14, Porphyromonas, Actinomyces, Alloprevotella, Rothia, Fusobacterium, Oribacterium, Lachnoanaerobaculum, Campylobacter, Kingella, Lautropia, Cardiobacterium, Peptostreptococcus, Catonella, Mogibacterium, F0058, Bergeyella, Capnocytophagia, Granulicatella,* and *Treponema_2*. Phylogenetic relationship of the top 20% most prevalent taxa is demonstrated in **Figure 7.4**.

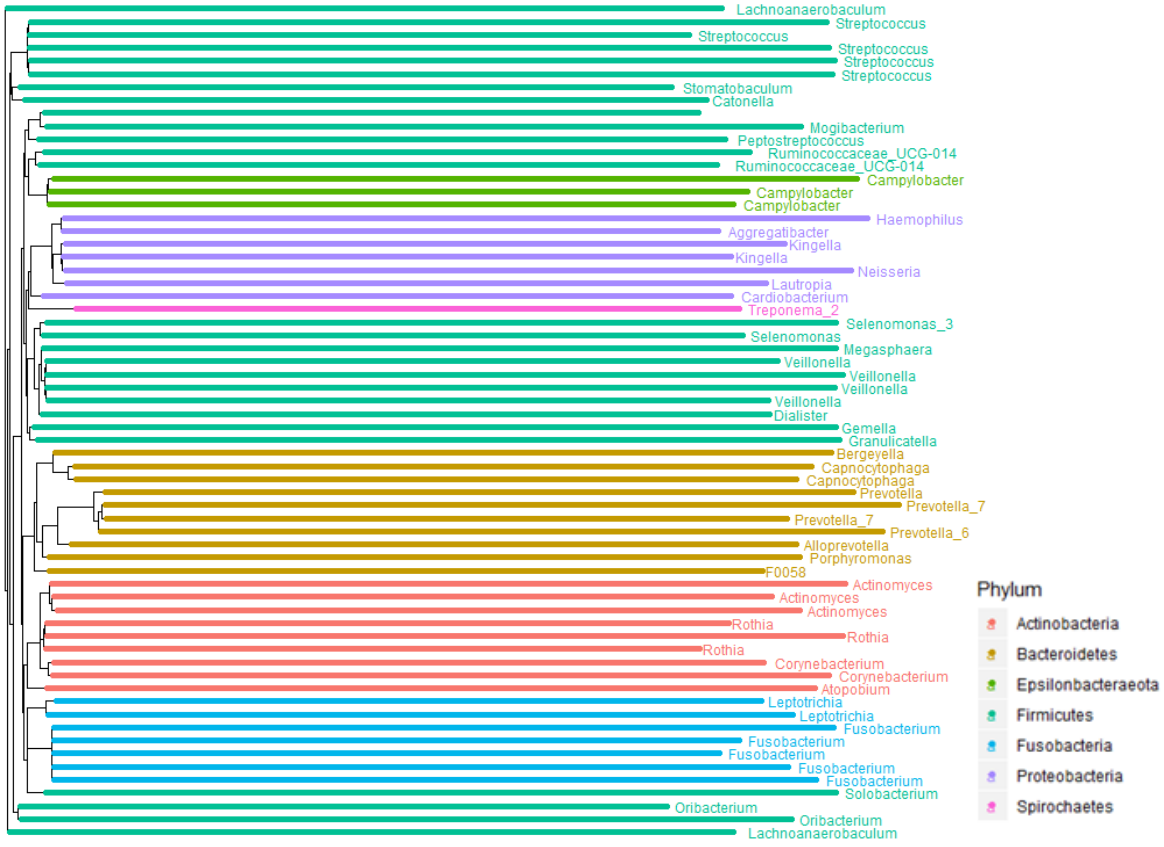


Figure 7.4 Phylogenetic tree of ASVs within more than 80% saliva samples
Tip labels show genus assignment of ASVs, and phylum is denoted by colour.

7.4.2 Characterisation of the saliva microbiota: Alpha diversity

Alpha diversity of the saliva microbiota was captured using six measures: Observed ASVs, abundance-based coverage (ACE), Shannon index, Simpson index, Inverse Simpson, and Fisher's alpha parameter (**Figure 7.5**).

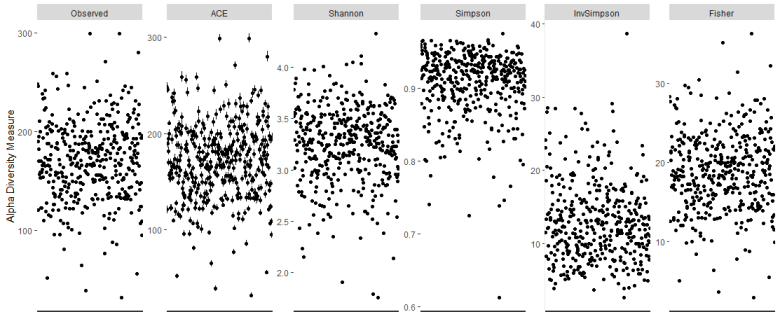


Figure 7.5 Alpha diversity of the saliva microbiota

7.4.3 Characterisation of the oral microbiota: beta diversity

Beta diversity was captured using Bray Curtis dissimilarity. On ordination, no clear association with age, frailty, fasting status of participants, or sample storage time was demonstrated (Figure 7.6). The statistical association between beta diversity and age, sex, BMI, diet, frailty, participant fasting status and sample storage time is presented below in Section 7.5.7.

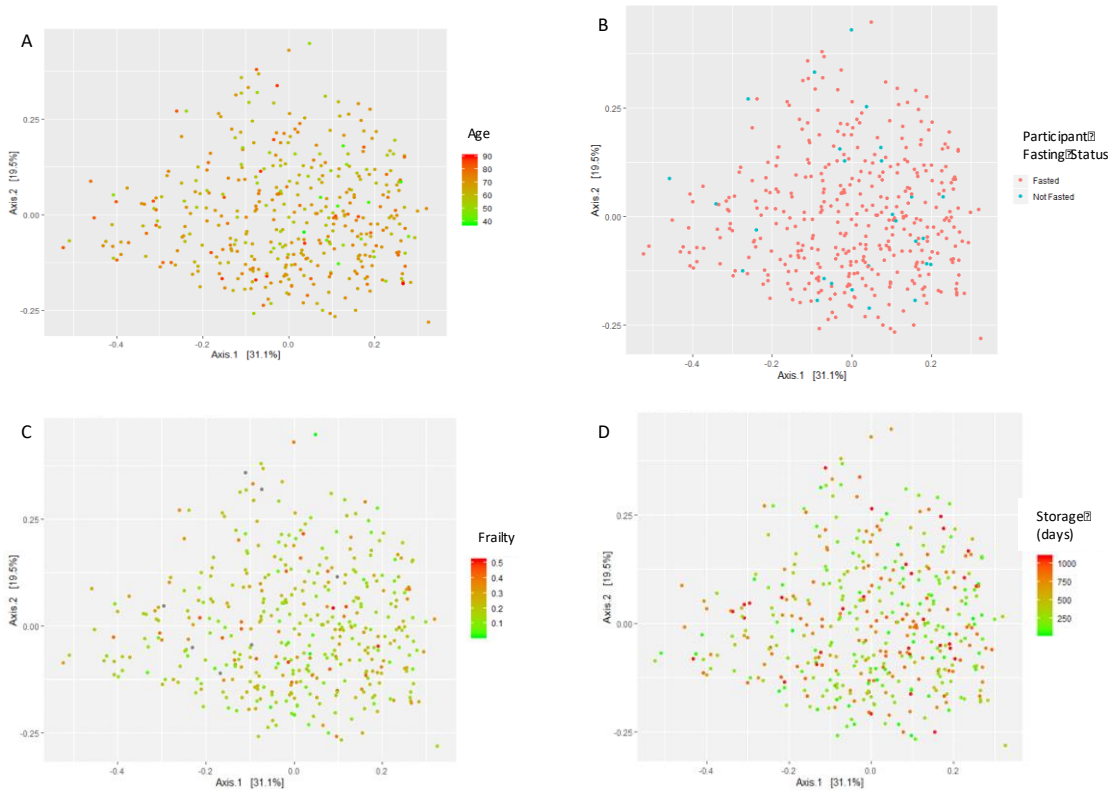


Figure 7.6 PCoA ordination of Bray Curtis dissimilarity of the saliva microbiota
Each data point represents a sample, coloured to reflect **A**) participants' age, **B**) fasting status of participants, **C**) general health as captured using the Rockwood frailty index, and **D**) Freezer storage time of samples in days as days.

7.4.4 Heritability estimates for taxa of the saliva microbiota

The saliva microbiota demonstrated substantial heritability (**Figure 7.7**). Within ASVs present in more than 15% of 688 samples, and therefore within the saliva microbiota of at least 100 participants, there were 110 taxa which were significantly heritable. The most heritable ASV was *Veillonella atypica* (44%, 95% CI 9 - 79), and the second most heritable ASV was *Campylobacter concisus* (43%, 95% CI 32 – 54). Heritability was not dominated by any phyla, however the most heritable phyla were Firmicutes and Bacteroides.

Figure 7.7 Heritability estimates for the saliva microbiota

Estimates of heritability with 95% for ASVs, present in more than 15% of samples (n=688), adjusting for age, sex, sample storage time and sequencing depth. Significantly heritable taxa (n = 110) are shown.

7.4.5 Association of the saliva microbiota with genetic risk of RA

TwinsUK participants (described in Section 7.2.1) comprised 407 adult twin volunteers, the majority of whom were female. An inclusion criterion of having been previously genotyped, and exclusion criteria of I) RA diagnosis; II) twin sibling with diagnosis of RA, were applied. Within saliva samples from TwinsUK participants, abundance of core taxa was associated with genetic risk for RA. There was no difference evident in alpha or beta diversity in association with the PRS.

7.4.6 Polygenic Risk Score

The RA polygenic risk score is described in Chapter 5. Briefly, the polygenic risk score was normally distributed within the TwinsUK cohort with RA cases and twin siblings excluded (Shapiro-Wilk normality test P= 0.24). The score ranged from 12.6 to 71.3; the median score was 42.63, the interquartile range 38.89 to 48.23. In 6,776 UK biobank participants, of which 2,686 had a HES diagnosis of RA, the score is predictive of RA. The standardised odds ratio for the PRS was 1.34 (P = 4.17e-8).

7.4.7 Alpha diversity

Genetic risk of RA was assessed for association with alpha diversity. There was no significant association between RA PRS and Shannon Index, Simpson index or Observed ASVs (Table 3).

	Shannon's diversity			Simpson's index			Observed ASVs		
	Estimate	P	Q	Estimate	P	Q	Estimate	P	Q
PRS	0.002714	0.65	0.731	0.005634	0.351	0.731	0.002714	0.659	0.743
BMI	0.007311	0.476	0.612	0.008038	0.439	0.612	0.007311	0.514	0.673
Frailty	-0.34179	0.457	0.612	-0.17067	0.714	0.871	-0.34179	0.544	0.721
HEI	0.000946	0.67	0.871	-9.19E-05	0.988	0.131	0.000946	0.772	0.978
Age	0.012649	0.029**	0.132	0.013366	0.022*	0.552	0.012649	0.029*	0.138
Male Sex	-0.25359	0.306	0.552	-0.25101	0.317	0.158	-0.25359	0.206	0.453
Fasting Status: Not Fasted	-0.39976	0.053	0.158	-0.47654	0.023*	0.552	-0.39976	0.057	0.158
Storage Time	-0.00016	0.299	0.552	-8.50E-05	0.58	0.701	-0.00016	0.289	0.53
Sequencing Depth	-1.31E-05	2.74E-07***	1.8E-0.6***	-1.09E-5	2.21E-05***	1.8E-07***	-1.31E-05	2.74E-07***	1.8E-0.6***

Table 7.2 Association of alpha diversity of the saliva microbiota with participant characteristics and sample technical factors

Using linear mixed-effects models, participant characteristics assessed were: Sex, Age, diet (HEI), general health (Frailty), BMI, and genetic risk of RA (PRS); technical factors assessed were storage time of samples within the freezer, and sample sequencing depth. Q value denotes adjustment for multiple testing of P values by application of the false discovery rate (FDR). Age was associated with all 3 measures of alpha diversity at nominal significance. Fasting status of participants was nominally associated with Simpson's index. Sequencing depth of samples was significantly associated with all three measures of alpha diversity.

Correlation between variables is shown in Table 7.3 A significant association was demonstrated between frailty and fasting status, whilst association between all other variables were non-significant.

	PRS	BMI	Frailty	HEI	Age	Male Sex
BMI	0.075					
Frailty	0.025	-0.114				
HEI	0.039	0.079	0.002			
Age	-0.007	0.059	0.011	0.065		
Male Sex	0.035	0.027	-0.034	-0.038	-0.135	
Fasted	-0.012	-0.090	-0.11	0.053	-0.083	0.013

Table 7.3 Rho coefficients of correlation between participant characteristics

Of all factors, only fasting status and frailty were significantly correlated with each other (p = 0.037).

7.4.8 Beta diversity

Beta diversity of oral microbiome samples was assessed. There was no association with RA PRS (**Figure 7.8;** PERMANOVA P = 0.387), however beta diversity was significantly associated with age (P = 0.001).

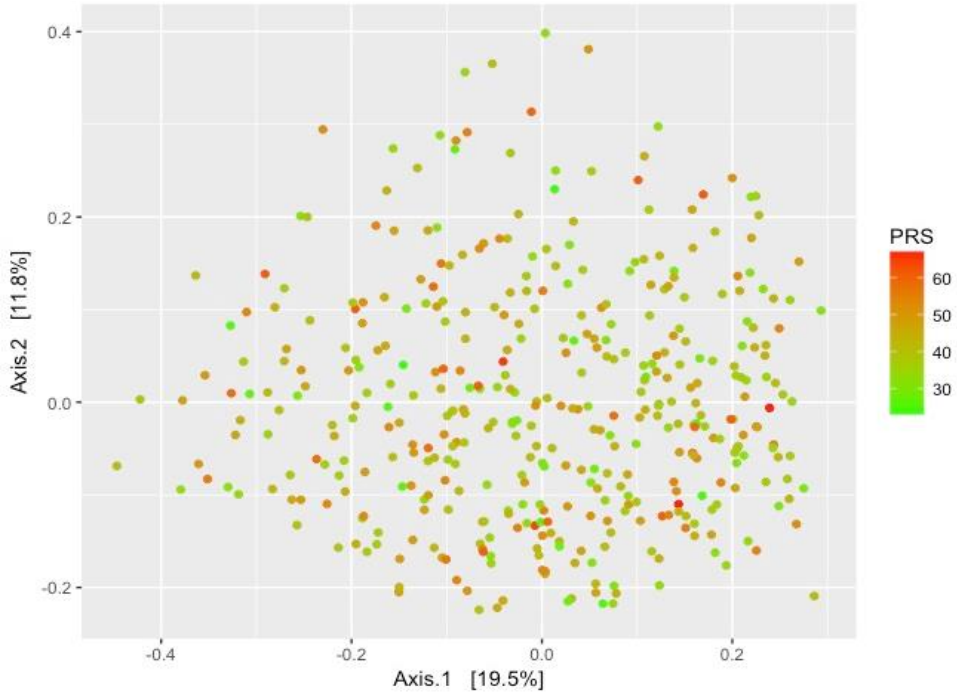


Figure 7.8 PCoA ordination of Bray Curtis beta diversity in relation to RA PRS
Each point represents a participant's sample. Samples from participants with a higher RA PRS are coloured in red, whilst those with a lower PRS are coloured in green (n=407).

	P value	Q value
PRS	0.304	0.447
Age	0.001***	0.008**
Sex	0.624	0.624
Relatedness	0.391	0.447
Storage Time	0.065	0.173
Fasting	0.337	0.447
BMI	0.024**	0.096
HEI	0.210	0.42

Table 7.4 PERMANOVA for Bray Curtis beta diversity of the saliva microbiota
Age was significantly associated with beta diversity after covarying for all other factors listed, whilst association with BMI reached nominal significance (n=407).

	P value	Q value
PRS	0.412	0.446
Age	0.003**	0.027*
Sex	0.617	0.617
Frailty	0.160	0.36
Relatedness	0.314	0.464
Storage Time	0.096	0.288
Fasting	0.402	0.464
BMI	0.085	0.288
HEI	0.233	0.419

Table 7.5 PERMANOVA for Bray Curtis beta diversity of the saliva microbiota
Age was significantly associated with beta diversity (q = 0.027) after controlling for all other factors. In this analysis, 7 participants without frailty data available were removed from the cohort (n = 400).

7.4.9 Taxonomic association between RA PRS and the saliva microbiota

Genetic risk of RA was negatively associated with 5 ASVs within the saliva microbiome (**Figure 7.9**). Of the taxon associations, 3 were within the 20% most abundant taxa (**Figure 7.4**).

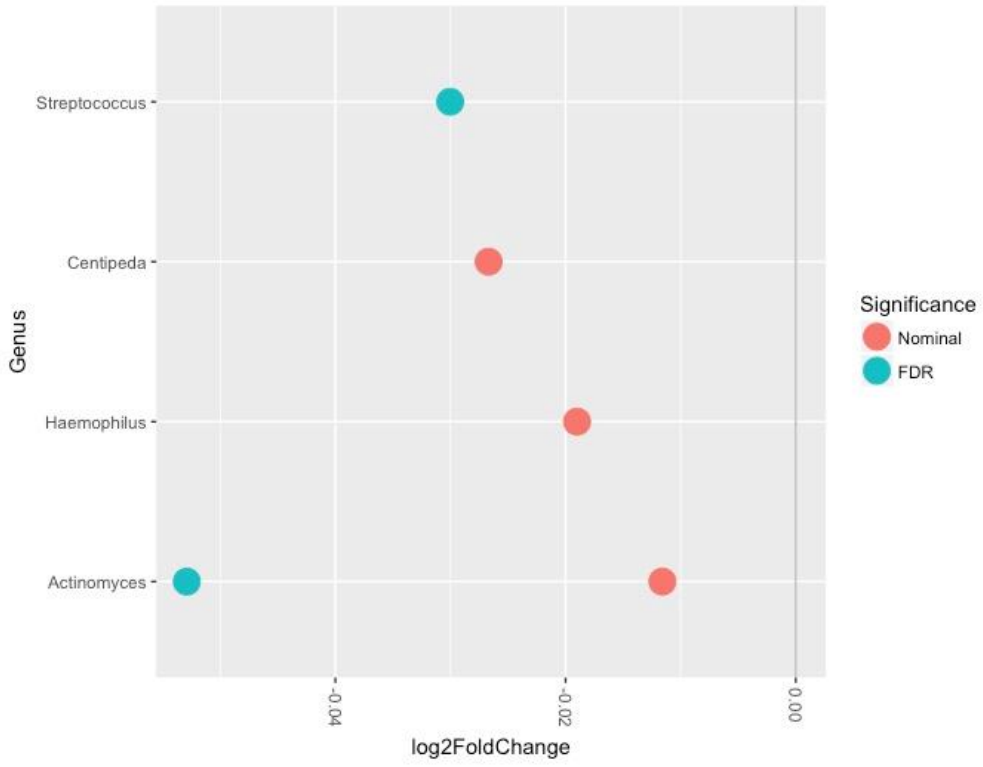


Figure 7.9. Difference in abundance of ASVs in relation to RA PRS

There were two taxa significantly associated with PRS after applying FDR to adjust for multiple testing: *Actinomyces* (Q = 0.03), and *Streptococcus* (Q = 0.03). There were 3 nominally significant ASV associations: *Centipeda* (p = 0.032), *Haemophilus* (p= 0.033) and *Actinomyces* (p = 0.006).

7.5 Discussion

7.5.1 Association of the oral microbiota with genetic risk of RA

In an RA unaffected sample of TwinsUK participants, genetic risk of RA was associated with composition of the saliva microbiota. The demonstration of association of saliva microbiota with genetic risk of RA is novel, and is consistent with an influence of RA risk genes with the oral microbiota; the study design excludes the possibility that the taxon associations are secondary to the physiology of established disease. Taxa identified as

associated with RA in prior studies, some of which comprise core features of the saliva microbiota, were associated with genetic risk for RA.

There were four taxa associated with RA genetic risk (Figure 7.9), all of which were significantly heritable (Figure 7.7). This is interesting as a stand-alone finding. Additionally, of the four taxa which were associated with genetic risk of RA, three have been previously associated with RA: *Haemophilus, Streptococcus* and *Actonomyces*. There is therefore potential for the taxa associations to be aetiologically relevant. Of these three, *Haemophilus* demonstrated the same direction of effect as prior studies of RA. For the other two taxa associations, the direction of effect is less clear: *Streptococcus* and *Actinomyces* have demonstrated conflicting associations with both positive and inverse associations with RA in prior studies (Zhang et al. 2015, Chen et al. 2018, Li, Li, et al. 2018).

Of the taxa demonstrated to associate with genetic risk of RA, *Haemophilus* was the most convincing both with regards to robust association with RA, and aetiological relevance, based on current evidence. In a large study by Zhang *et al.* in which metagenomics techniques were used to compare the oral microbiota of RA patients and health control participants, *Haemophilus* and *Streptococcus* in the saliva microbiome were inversely associated with RA (Zhang et al. 2015). A study by Chen *et al.* replicated these findings, also demonstrating an inverse association of *Haemophilus* and *Streptococcus* with RA. Epidemiological evidence is important, but perhaps more convincingly with regards to functional relevance, *Haemophilus* was associated with clinical parameters of RA; abundance of *Haemophilus* inversely correlated with CRP, RF and ACPA (Chen et al. 2018). There is therefore some support for an aetiological link between genetic risk of RA and the saliva microbiota.

The context of the association with *Streptococcus* is less clear. There are conflicting findings as to the direction of association with RA in prior studies (Scher et al. 2012, Zhang et al. 2015). *Streptococcus* is also associated with periodontal disease, however the direction of effect most often reported is positive. For example, higher abundance of *Streptococcus*

mutans in saliva and plaque is found in periodontal disease patients, and correlation demonstrated with clinical parameters (Dani et al. 2016). A probable explanation is that the *Streptococcus* in the present study is a different species or strain.

For the third taxon association, *Actinomyces*, there is some evidence for implication in RA aetiology. However, it is only directionally consistent with associations demonstrated for periodontal disease. A lower abundance has previously been associated with periodontal disease (Wang, Qi, et al. 2013, Li, Li, et al. 2018). In the study by Zhang *et al.* a higher abundance of *Actinomyces* in the saliva microbiota of RA patients compared to healthy controls was reported, in contrast with the present study (Zhang et al. 2015). The relevance of the periodontal disease association with *Actinomyces* is tentative; Given that there is limited understanding of shared genetic loci between RA and periodontal disease at present, the likelihood of RA genetic risk influencing abundance of taxa which are characteristic of periodontal disease is unknown, though it is possible. It is difficult to draw conclusions without ascertainment at the species level and below. However this taxon has been identified as having pathophysiologic relevance in periodontal disease, having been identified in prior work as an important species having influence on bacterial community structure in subgingival plaque biofilm (Li, Li, et al. 2018).

An association was not demonstrated between genetic risk of RA and *P.gingivalis.* As *P.gingivalis* has been found in higher abundance in periodontal disease, which is also linked to SE risk alleles, and has been associated with RA pathology, it was potentially feasible that *P.gingivalis* would positively associate with genetic risk of RA. However the results of the present study are not conclusive, as *P.gingivalis* did not pass the abundance filtering threshold and was therefore not included in the taxon association analysis; *P. gingivalis* was not present within more than 40% of the saliva samples. Three other *Porphyromonas* species passed the filtering threshold: *Porphyromonas Pasteri*, *Porphyromonas endodontalis*, and *Porphyromonas catoniae*. It is interesting to note that the ASV methods allowed for identification and annotation of this taxon to species level; This would not have been possible using OTU based methods, with which accurate determination is not possible

beyond genus level. An ASV assigned to *Porphyromonas* was demonstrated to be 30% heritable, however here species assignment was not possible and it is unknown whether this ASV is *P. gingivalis.*

There may be a link between *P.gingivalis, Actinomyces* and *Streptococcus.* In a study by Colombo *et al.,* it was demonstrated that *P.gingivalis* outer membrane vesicles facilitate coaggregation of *Actinomyces* and *Streptococcus*. This suggests that there is a synergistic ecological community interaction, in support of the pathological biofilm hypothesis of these taxa (Colombo et al. 2016).

In accordance with prior studies of the oral microbiota in RA (Scher et al. 2012, Zhang et al. 2015), no association between genetic risk of RA with alpha diversity was demonstrated. Directionally, the (non-significant) association was positive. Periodontal disease has been associated with higher alpha diversity compared to healthy controls, however, the implications of this are unclear. The oral microbiota do not follow a similar pattern to the gut microbiota, in which a lower alpha diversity is generally associated with a diseased state (Jackson et al. 2018a). This does not necessarily indicate that higher alpha diversity is beneficial in the gut microbiota, or conversely that a lower alpha diversity could lead to poorer health outcomes. In the oral microbiota, there is a varied pattern of association; higher rather than lower alpha diversity is sometimes associated with disease (Takeshita et al. 2016). Replication across different cohorts, and consideration of methodological differences between studies is required to understand alpha diversity of the oral microbiota.

Overall, in a relatively small sample size, a signal was evident for PRS association with the saliva microbiome. Replication in another dataset would make the finding more robust. Potentially the association of the saliva microbiota with RA PRS also captures wider association with ill health. In support of this, RA PRS was positively, but non-significantly, correlated with lower general health (higher frailty index; Table 7.4). However, neither

frailty nor PRS were demonstrated to associate with alpha diversity of the saliva microbiota in the present study.

It is interesting to consider the contrast in taxa microbiota association with the PRS in saliva versus intestine. The association of RA PRS with the gut microbiota (Chapter 5) demonstrated a larger effect size, and entirely different taxon associations that with the saliva microbiota. In addition, the association with alpha diversity demonstrated opposite directionality, of positive in saliva and inverse in the intestine, though in both instances significance was not reached and conclusions cannot be robustly drawn. In the gut microbiota, a key determinate of microbiota composition is inflammatory milieu, and genetic risk of RA is determined by immune pathways. Speculatively, in the saliva microbiota immune factors may be less influential.

7.5.2 Characterisation of the saliva microbiota

The micro-environments of the oral cavity and associated commensal microbiota of the respective niches are described in Chapter 1. However, there is an ongoing research effort to understand the composition of the oral microbiota. In particular, there are limited prior studies which characterise the oral microbiota in relation to potential influencing factors. I therefore undertook heritability analysis and characterisation of the saliva microbiota in relation to methodological factors (e.g sample storage time) and participant characteristics.

To my knowledge the heritability of the adult oral microbiome has not been previously reported. In the present study the saliva microbiome demonstrated substantial heritability: of taxa (ASVs) present in at least 100 participants, 110 were significantly heritable, with heritability ranging from 44 % - 13 % (Figure 7.7). The most heritable phylum was Firmicutes.

For saliva microbiota diversity, of potential influences considered - age, sex, BMI, diet, general health (frailty), relatedness, sample storage time, and fasting prior to provision of sample - association was demonstrated with three factors: age, BMI and fasting status of participants. Only the association with age remained significant after adjustment for multiple testing. Age was associated with both alpha and beta diversity. Fasting status was associated with alpha diversity at nominal significance, and BMI was associated with beta diversity also at nominal significance.

In the present study, age was therefore the factor most strongly associated with saliva microbiota composition, having association with both alpha and beta diversity which remained even after adjustment for health deficit (frailty) and multiple testing. Regarding alpha diversity, age was significantly and positively associated with alpha diversity in all three measures; higher age was associated with increased alpha diversity as demonstrated by Shannon diversity ($p = 0.029$), Simpson's diversity ($p = 0.022$), and observed ASVs ($p = 0.029$), Table 7.2). Analysis of beta diversity (Bray Curtis) showed a significant association with age ($Q = 0.027$; Table 7.5) even after adjusting for general health (frailty).

The second strongest association with beta diversity (Bray Curtis dissimilarity) was BMI, however only nominal significance was demonstrated ($P = 0.024$), the association did not remain after adjustment for multiple testing ($Q = 0.096$). After adjustment for frailty in the subset analysis of participants for whom frailty data were available, no association between BMI and beta diversity was demonstrated.

Simpson's index measure of alpha diversity was also nominally associated with fasting status ($p = 0.022$), with an inverse association of non-fasted samples, For Shannon's alpha diversity, the association was approaching nominal significance ($p = 0.052$). This indicative result with a small sample size suggests that the true association is stronger, but the present study was underpowered to detect it: in the TwinsUK sample, only 7% (28) of participants had not fasted. Further work including a larger sample of non-fasted participants will be required to confirm the relationship with alpha diversity. Non-fasting

participants are potentially more likely to be have type 2 diabetes, which should also be taken into account for association with the saliva microbiota (Sabharwal et al. 2019). However, these results indicate that not fasting may influence microbiota composition to reflect recent consumption, which could introduce study bias. In general, studies of the oral microbiota do not take account of fasting prior to provision of a saliva or other oral microbiota sample (Zhang et al. 2015), but this may be important moving forward.

In a study by Wu *et al.* which characterised the salivary microbiome of 62 younger adults in relation to obesity, an association of BMI was demonstrated with both alpha diversity (Chao1, $P < 0.01$; Shannon diversity, $P < 0.05$) and beta diversity (unweighted UniFrac, $P = 0.001$) (Wu et al. 2018). The study design accounted for age, sex and oral hygiene, which were not statistically significant between cases and controls. In TwinsUK, BMI was nominally associated with beta diversity as calculated using Bray Curtis dissimilarity prior to adjustment for health deficit (frailty), but no association was demonstrated after adjustment. No association was observed between BMI and alpha diversity. Wu *et al.* used targeting of the V3 variable region, which could potentially account for the difference in results if there is higher sensitivity for taxa which associate with obesity. However the nominal association with composition prior to adjustment for frailty in TwinsUK suggests that un-measured higher multi-morbidity in the obese group in the study by Wu *et al.* could account for some of the association they demonstrated between BMI and the saliva microbiota.

A study of the saliva microbiota by Ogawa *et al.* included consideration of frailty. Of participants, 16 lived in a nursing home and 15 lived within the general community (Ogawa et al. 2018). Those who were nursing home dwellers had a mean age of 87, whereas those who lived within the community had a mean age of 84. Participants who lived in a nursing home had the inclusion criterion applied of recommendation by a medical doctor to live in a nursing home due to frailty. The authors therefore classified the nursing home group as frail and the independent community dwellers as non-frail. Ogawa *et al.* demonstrated a significant inverse association of alpha diversity of the saliva microbiota with frailty. This is

directionally consistent with the non-significant trend demonstrated in the present study. In the study by Ogawa *et al.* participant groups were selected to have a large discrepancy in frailty: They investigated the oral microbiota of elderly adults of comparable age living either in the community or in a care home setting, with the latter comprising the frail participants. They accounted for age, BMI and dental health, but included no adjustment for diet or sex. The difference in diet between the non - frail and frail groups is likely to be substantial, being home and institution dwelling participants, respectively. This may therefore represent a major limitation of the study. Nonetheless, further work is needed to understand the link between diet and oral microbiota composition. Sex distribution was different between groups: 3 out of 15 in the frail group and 9 out of 16 in the non-frail group were male. In the present study there may be less variation in the frailty of participants, which comprise TwinsUK community-dwelling volunteers and this may explain the discrepancy in results. However this is difficult to demonstrate, as Ogawa *et al.* did not quantitatively measure frailty as here.

A recent population study by Burcham *et al.,* considered factors which influence the oral microbiota in both adults and children, with separate analysis of each group (Burcham et al. 2020). Oral samples were obtained using buccal swabs of the teeth, tongue, cheeks and gums. Therefore, multiple oral sites were included, in addition to saliva. The source of saliva microbiota is predominantly the biofilm on the dorsum of the tongue, however all other oral microbial niches contribute taxa (Davenport 2017). The buccal swabs used here are therefore relevant to the saliva microbiota, but may not be directly comparable. Burcham *et al.* included 172 adults aged between 20 and 57, median age 34. They did not consider age within the adult or child groups, but on comparison of both groups they demonstrated significantly lower Shannon diversity in adult versus child groups. Within both groups they considered weight status, sex, prescription of antibiotics in the last 6 months and oral hygiene habits (visits to the dentist for professional descaling). They showed that in adults only oral hygiene habits were significantly associated with beta diversity. There was no association between weight status, sex or ingestion of antibiotics in last 6 months. However, youth oral microbiome beta diversity was associated with both sex and weight

(Burcham et al. 2020). In the present study, oral hygiene habits were not accounted for, but could potentially be captured in part by age, with a pattern of declining oral hygiene in older age.

Diet as captured using the HEI was not significantly associated with alpha diversity of the saliva microbiota. This is in contrast with findings for the gut microbiota in TwinsUK. We have shown that diet is significantly associated with Shannon diversity, Simpson's index and observed OTUs (Bowyer et al. 2018). This is surprising as it may be expected for food substrates in the oral cavity to allow some bacteria to flourish relative to others, dependent on adaptation to metabolic factors and environmental niche pH.

In TwinsUK saliva samples, the most abundant phylum within the saliva microbiota was demonstrated to be Proteobacteria. This was constant when considering highly prevalent taxa, found in more than 80% of samples (**Figure 7.3**), ASVs present in more than 40% of samples (**Figure 7.2**), and when considering the full dataset of all ASVs (**Figure 7.1**). In the present study, employing sequencing of the V4 16SrRNA gene variable region, the taxonomic composition demonstrated is in contrast with prior studies, the majority of which targeted the V2 variable region, and show Bacteroidetes to be the dominant phylum. (Tsuda et al. 2015, Gomez et al. 2017, Ogawa et al. 2018, Murugesan et al. 2020).

Murugesan *et al.* undertook a large recent study of the saliva microbiota of 997 younger adults with mean age of 38, using targeting of the V2 variable region, demonstrated that the dominant phylum of the saliva microbiota was Bacteroidetes (Murugesan et al. 2020). Proteobacteria were the third most dominant phylum. This finding was replicated in studies by Ogawa *et al.*, *Gomez et al.*, and Tsuda *et al.* (Tsuda et al. 2015, Gomez et al. 2017, Ogawa et al. 2018).

Ogawa *et al.* undertook a small study of 31 elderly adults. Within community dwelling older adults, in contract to our results, the saliva microbiota were dominated at the phylum level by Bacteroidetes. Proteobacteria was the third most abundant phylum, after

Firmicutes. There are important methodological differences compared to the present study, which are likely to account for the taxonomic discrepancy. Ogawa *et al.* used primers which target the V2 hypervariable region, generated de-novo OTUs and assigned taxonomy using the Greengenes database. In the present study, the V4 region was targeted, ASVs were generated and taxonomy was assigned using the Silva database. They also required their participants to refrain from food, drink, smoking and use of toothpaste for 2 hours prior to saliva collection. In the present study, participants were asked to fast and refrained from smoking or chewing gum for at least 6 hours, whereas there was no stipulation about toothpaste. Ogawa *et al.* undertook DNA extraction, PCR and sequencing immediately, whereas in the present study samples were frozen prior to these steps. The primary factor mediating the taxonomic discrepancy between both studies is likely to be choice of variable region targeted by the PCR primers.

Tsuda *et al.* undertook a study of the saliva microbiota of 44 adults who had fasted overnight, using pyrosequencing after PCR targeting of the V1-V2 hypervariable region (Tsuda et al. 2015). The dominant phylum was Firmicutes, whilst Bacteroidetes, Actinobacteria and Proteobacteria were the second, third and fourth most abundant, respectively. Another study, by Lundmark *et al.* evaluated the saliva microbiota of 114 adults, using targeting of the V3-V4 hypervariable region (Lundmark et al. 2019). They demonstrated that the most abundant phyla in order of dominance were Firmicutes, Bacteroidetes, and Proteobacteria.

There are few studies of the saliva microbiota which use targeting of the V4 variable region, however in a small study of 20 adult participants, using this method, in accordance with our study, Proteobacteria was the most dominant phylum, whilst Firmicutes was the second most dominant (Yu et al. 2020).

In the study by Gomez *et al.* the dominant phylum of the subgingival plaque of children aged 5 to 11 is Firmicutes (Gomez et al. 2017). There are three key methodological differences compared to the present study: age, oral site and variable region targeted.

Gomez and colleagues used primers which target the V2 variable region, whereas in the present study V4 was targeted.

These high phylogenetic level taxon discrepancies between studies which employ targeting of different variable regions of the 16S rRNA gene indicate that primer selection is an important factor having influence on downstream findings. The primers used in this study, 515F and 806R, were initially used by the Earth Microbiome Project, and have since been modified for use with the Illumina platform (Caporaso et al. 2011). This modification rectified the previously held bias against *Crenarchaeota* and *Thaumarchaeota* (Hugerth et al. 2014). These primers have been demonstrated to perform well for characterising the gut microbiota, however efficacy for characterisation of the oral microbiota requires further investigation. A recent study showed a discrepancy in performance when applied to human gut versus skin microbiota (Wasimuddin et al. 2019). Most studies of the saliva microbiota have employed primers which target the V2 variable region. This region demonstrated higher resolution for *Streptococcus,* the most abundant genus in the oral cavity (Wade and Prosdocimi 2020).The performance of the primers used in this study with regards to the saliva microbiota requires further investigation.

The work here has demonstrated overall composition of the microbiota and association with potential influencing factors. Of all factors, age was had highest importance with regards to microbiota diversity, whilst an influence of BMI and fasting status of participants was also suggested. It is also important to note that for all factors there may be association of specific taxa and ecological communities within the saliva microbiota; further work is required to demonstrate these.

7.5.3 Consideration of strength and limitations of the study

Some key strengths of the study lie in the technical aspects of sample processing. DNA extraction and sequencing was performed by the same person, samples from twin pairs were separated on the DNA extraction plates and all samples were included in two lanes of the same sequencing run. The inclusion of blank reagent only samples allowed for removal of potential contaminant ASVs from the dataset, sometimes referred to as the "kitome". Consideration of contaminants is particularly important for lower biomass samples such as saliva.

However there were also limitations to the work which are important to consider. We used targeting of the V4 region which, as discussed above, has lower affinity for key taxa of the oral microbiota, and particularly *Streptococcus.* We used sampling of the saliva microbiota, which derive predominantly from oral mucosal surfaces. However, the microbiota of other sites, and particularly subgingival plaque may be more pathologically relevant. The taxon association analysis of RA PRS and the oral microbiota did not take account of some potential influencing factors, the most important of which may be smoking, fasting prior to sample collection, oral health habits. In the study protocol, there was no stipulation as to oral health care just prior to providing the sample. Quite feasibly, participants may have deviated from their usual oral hygiene habits in anticipation of providing a saliva sample; this was not accounted for in our study. Periodontal disease was not accounted for (discussed below). There is potential bias introduced by the features of the TwinsUK cohort from which the study participants were sampled. The volunteer cohort population is predominantly female, white and of higher affluence. The limitation of the sex ratio of the cohort is offset as RA is more common in females. Ethnicity is a limitation as RA is more common in Black people and I was unable to explore association in different ethnicities as the study sample was 99% Caucasian. RA is more prevalent in areas of socioeconomic depravation, however understanding is that this is secondary to lifestyle factors such as smoking and diet, and this therefore should not bias a study of genetic factors.

The present study did not include consideration of periodontal disease, this would be informative to explore; periodontal disease - an established risk factor for RA - and RA, are associated with overlapping taxon associations in the oral microbiota, and both conditions share some common genetic risk factors. Periodontal disease has been historically associated with a group of bacteria known as the 'red complex', which comprises *Porphyromonas gingivalis*, *Treponema denticola* and *Bacteroides forsythus,* all of which were reported as particularly associated with clinical measures of periodontal disease – dental pocket depth and gingival bleeding on probing (Socransky et al. 1998). It has since been advanced that wider microbiota associations or specific gene combinations thereof converge to shape and stabilise a disease provoking microbiota. Disease development could be secondary to presence of a pathogenic oral microbiota in the context of other pre-disposing risk factors and genetic propensity for inappropriate immune response.

Both periodontal disease and RA are pre-disposed by the HLA-DR Shared Epitope (SE) and in addition both conditions share SNPs and epigenetic modifications relating to cytokine genes, and in particular expression of IL-6, which is of key importance in both RA and periodontal disease (Bingham and Moni 2013). However, it is important to note that the strength of association of the SE with periodontal disease is much weaker than with RA. There may be wider overlap in the genetic aetiology of RA and periodontal disease; further work is required to demonstrate this.

7.5.3 Considerations for future work

Primarily, replication in another cohort is needed. To extend the study, there are additional analyses that would be informative: In particular, further measures of alpha and beta diversity, and analysis of ecological community association. For the latter, the method applied in Chapter 6 of balances would be appropriate. Characterisation of the oral microbiota in the wider dataset (N = 707) should be undertaken, as in the present study

those who had not been genotyped were excluded. It would be informative to repeat the present study using samples of different oral niches, for example sub-gingival plaque, dorsum of tongue, teeth, and buccal mucosa. Previous associations have centred around plaque on the tongue, teeth and in sub-gingival pockets (Eriksson et al. 2019).

Given the identified differences in salivary microbiome composition observed in this study with fasting status, further sensitivity analyses could be conducted to check robustness of the associations excluding the 7% of individuals who were non-fasted. This approach could also be considered for the heritability analysis. Smoking status is an additional factor which has not been considered in this analysis, although rates of smoking are relatively low in this cohort, at around 8 % (Bowyer et al. 2018).

Further work is needed to investigate periodontal disease as it is suggested to be on the causal pathway of RA, with overlapping taxon associations with the oral microbiota. Understanding of the genetic aetiology is not sufficient to calculate a polygenic risk score, therefore a study of periodontal disease using similar methods to the present study would not be viable. However, quantifying the shared genetic aetiology between RA and periodontal disease would provide much needed clarity of the association; this could be achieved using bivariate classical twin modelling. To account for shared taxon associations between periodontal disease and RA it may be informative to extend the study by taking a nested model approach for periodontal disease. To investigate taxon associations with periodontal disease, taxon differential abundance analysis should be undertaken in which periodontal disease is considered as a response variable, as conducted for RA PRS in the present chapter and RA in Chapter 5.

In order to contextualise the findings of this study, and to extend understanding of the general research area, it will be important for studies to continue to characterise the oral microbiota of RA patients. There are few studies completed to date, with discrepancy in findings. As above, such studies would benefit from taking into account different oral niches. These studies should also consider association of taxa with clinical parameters such

as ACPA and DAS28, and also immune parameters such as antibody titre in relation to specific taxa, and inflammatory immune cell profile. Functional work should be undertaken, characterising the mechanistic link between oral taxa and RA. For example, use of *in vitro* methods including organoids, and *in vivo* methods including animal models.

7.6 Conclusions

The saliva microbiota of adults demonstrated substantial heritability. Composition of the saliva microbiota is associated with age even after adjustment for general health. In a non-RA cohort, genetic risk of RA was associated with four taxa of the saliva microbiota. Whilst further work is required to offer a robust conclusion, this study tentatively supports the hypothesis that association of the oral microbiota with RA is secondary in part to host genetic factors.

Addendum

Following completion of the thesis, the study to characterise the saliva microbiota has been expanded to include a larger sample size and an additional phenotype, periodontal disease. For brevity, this work has been included in an additional appendix - Appendix E.

Chapter 8

General Discussion

The overarching aim of this thesis is to contribute to understanding of the link between the commensal microbiota and RA aetiology. In this final chapter, I consider the implications of the thesis findings in their entirety and evaluate whether the thesis aims have been addressed. For clarity, I include an overview of thesis findings, prior to a more detailed discussion in the context of the thesis aims and the wider research area. I consider limitations of the methods, and finally I suggest how the work could be developed moving forward for the furtherment of understanding of the link between the commensal microbiota and RA.

8.1 Summary of thesis findings

Within this thesis I have applied novel approaches to elucidate the link between the commensal microbiota and RA aetiology. I first considered 16S sequence pre-processing methods, and curated the TwinsUK ASV dataset, presented in Chapter 3. Heritability estimates for the ASVs were presented in the context of demonstrating efficacy of methods compared to those used previously. Whilst not conclusive in this regard, the analysis supported convincing prior studies benchmarking the method.

In Chapter 5 I investigated the association of genetic risk of RA with the intestinal microbiota. I demonstrated that a group of *Prevotella spp., Prevotella_7,* were associated with genetic risk of RA in participants without the disease. As part of our collaboration with a Swiss group which is led by Axel Finckh and Till Robin Lesker, I found within the

SCREEN-RA cohort a higher abundance *Prevotella_9,* (predicted *Prevotella copri*), in intestinal microbiota of participants with pre-clinical RA (having signs and symptoms of RA, but not fulfilling the clinical criteria for diagnosis), compared to healthy control participants. The ASV methods afforded demonstration that *Prevotella_7* did not include *Prevotella copri,* however the ecological community- based analysis in this chapter provided evidence for an interaction between *Prevotella copri* and the *Prevotella_7*, in which abundance is interdependent between these taxa. This association of *Prevotella _7* with genetic risk of RA (PRS) in Twins UK was validated the SCREEN-RA cohort, in which abundance of *Prevotella_7* was associated with presence of shared epitope RA risk alleles. Further supporting this work, the heritability analysis in Chapter 3 demonstrated that *Prevotella_7* was significantly heritable, whilst *Prevotella copri* was not.

In Chapter 6 I took an opposing approach of investigating association of the gut microbiota with RA, whilst controlling for host genetic factors, within a discordant twin study design. Here, I investigated differences in the intestinal microbiota of established (treated) RA patients, compared to a control group comprised of RA affected participants' unaffected twin siblings. Within the sample of monozygotic twin pairs, there were four taxa which demonstrated a higher abundance in RA affected versus unaffected twins. The strongest taxon association, *Bacteroides,* is a dominant taxon associated with modulation of inflammation. *Bacteroides fragilis* has been previously demonstrated to have higher abundance in both healthy controls and treated RA patients, compared to new-onset (untreated) RA patients. However it was not possible to identify to species level the taxa associations of this study, and interpretation is therefore unclear: if the association is with *Bacteroides fragilis,* this potentially corroborates a prior study indicating that anti-inflammatory taxa are upregulated in the gut microbiota following RA treatment with DMARDs. The compositional approach of taxon balances demonstrated no significant differences in relationships between intestinal microbiota of RA affected versus RA unaffected twins.

In Chapter 7, I investigated the association of the saliva microbiota with genetic risk of RA.

As part of this I performed characterisation of the saliva microbiota in relation to a range of potentially important biological and technical factors. Whilst heritability of the adult oral microbiota has not been previously published, I demonstrated that there is substantial heritability of the saliva microbiota; 113 taxa of the saliva microbiota were significantly heritable. Age was associated with saliva microbiota diversity after controlling for other potential influences, including general health. The association of genetic risk of RA with the saliva microbiome comprised four taxa, providing some evidence of a potential link between host genetic factors and the saliva microbiome in RA; the literature evidence base surrounding these taxon associations and particularly the direction of association was less strong.

8.2 Consideration of findings with regards to thesis aims

The thesis aims are presented in Chapter 2. For clarity, these are also included here.

8.2.1 Overarching aim: Aim 1

1) *Contribute to understanding of the link between the gut commensal microbiota and RA.*

Understanding of the link between the commensal microbiota and rheumatoid arthritis is presently in the exploratory phase. It is presently subject to a substantial multicentre research effort. In completing this thesis, I have contributed several novel findings which will help bring clarity to the research area; these findings are discussed in the following sections for specific aims.

8.2.2 Specific aims: Aim 2

2) *Apply appropriate methods to the study of the commensal microbiota in RA.*

In this thesis I have applied many different methods to appropriately analyse microbiome data in relation to phenotypes, including genomic and clinical data. I have applied appropriate statistical methods, which take into account the specific considerations for microbiome data.

For example, ASV methods were published by Callahan *et al.* just prior to commencement of my PhD in 2017 (Callahan et al. 2016, 2017); using these novel methods I generated ASVs in TwinsUK data. The benefits of this approach are most clearly demonstrated by the results of the study in Chapter 5. When I performed the same analysis using OTUs, no associations were shown. Importantly, using ASVs I was able to identify *Prevotella* and also was able to taxonomically discriminate the *Prevotella* associations further than is possible with OTUs. This was aided by use of the SILVA database which has improved taxonomic assignment than previously used databases (e.g. Greengenes). Thus, the study in Chapter 5 demonstrated significant association between genetic risk of RA and *Prevotella_7* and *Prevotella_9* with genetic risk of RA and pre-clinical RA, respectively. Using this additional taxonomic classification, I was able to identify that the association was not actually with *Prevotella copri,* often the most abundant *Prevotella* species, and previously reported in association with RA, but with a separate more heritable *Prevotella* which is in balance with *Prevotella Copri*. Methods used thus afforded further interrogation of the *Prevotella* association to avoid outputting misleading results. The interesting results of this study are therefore attributed to the methods chosen.

8.2.3 Specific aims: Aim 3

3) Investigate the gut microbiota association with RA with regards to host genetic factors.

In **Chapter 5**, I demonstrated that there is association of RA host genetic factors with RA relevant taxa in the gut microbiota. The strongest taxon association with genetic risk of RA, *Prevotella_7*, and there was a suggested biological link with the primary RA taxon of interest, *Prevotella_copri.* Informatively, the heritability analysis of the ASVs which I presented in **Chapter 3** showed *Prevotella_7* to be significantly heritable, whereas *Prevotella_9* (predicted *Prevotella copri*) was not. There has been prior suggestion of a link between presence of Shared Epitope alleles and abundance of *Prevotella copri* (Scher et al. 2013), which this work lays into question.

As discussed in detail in **Chapter 2** and **Chapter 5**, there is a body of evidence indicating a functional link between *Prevotella copri* and the pathogenesis of RA. More recently the role of other *Prevotella* species have additionally been highlighted. With regards to Aim 3, this work is not considered conclusive. However, it supports the hypothesis that *Prevotella spp.* are involved in RA pathogenesis. To put the findings in context, it is likely that there are other factors which are more strongly associated with abundance of *Prevotella copri*. For example, it is established that diet is a mediator of *Prevotella* abundance. In TwinsUK, we have recorded dietary data and using this my colleagues have curated dietary indices. However as discussed in **Chapter 5**, there were missing data for a portion of the study sample, and when repeating the model in the subsample for whom dietary data was available, no association for *Prevotella* was shown. On further investigation, the participants for whom dietary data was unavailable had the samples with higher *Prevotella_7* and repeating the model in this subset yielded no association for *Prevotella_7.* It is a lower-prevalence taxon, and the excluded samples therefore affected the results. In the future I would like to explore different solutions such as imputation methods. However, in the absence of this, the convincing prior evidence for the role of diet (Gorvitovskaia et al.

2016) generates the hypothesis that diet is more influential than host genotype with regards to abundance of *Prevotella.* Indeed, there is recent suggestion that additionally, regardless of difference in abundance, *Prevotella spp.*, particularly *P.copri,* may function differently in the context of differing host diets. In a study by De Fillipis *et al.,* carbohydrate active (CAZ) enzymes were associated with a diet high in carbohydrate (De Filippis et al. 2019). De Fillipis *et al.* hypothesised that there is a different functionality of *P. copri* according to host diet, independent of *P.copri* abundance. As discussed in Chapter 5, there are also conflicting findings regarding pathogenicity, or conversely, beneficial status of *Prevotella copri*; there is evidence for both ends of the spectrum, spanning beyond the RA microbiome field. A potential explanation in addition to the CAZ enzyme evidence is that conflicting findings relate to different strains, with different functionalities. In support of this, four distinct clades have recently been identified, termed the "*Prevotella copri* complex" (Tett et al. 2019). In addition, the *Prevotella copri* genome contained multiple transposable elements, with potential to move between taxa (Moss et al. 2020). Overall *Prevotella copri* is quite an unusual and intriguing taxon. To understand it furthermore specificity in taxonomic identification is required as well as contextualisation in the wider ecological community and in relation to functions rather than presence or absence. For this metagenomics, and potentially metatranscriptomic methods will need to be deployed.

In Chapter 6 the converse of the above approach was taken: genotype was controlled for in a discordant twin study design. When potential influential factors of genotype, age, sex, BMI, diet, and early-life factors are controlled for between RA discordant MZ twins, significant taxon differences are still detectable. The largest difference in taxon abundance was an increase in *Bacteroides* in RA affected compared to control twins. It is difficult to draw robust conclusions, partly because prior studies have focused on early and as yet non-treated RA, and partly because the RA affected twins in our cohorts are taking RA medications which may affect their gut microbiota composition. However an unusual pattern within the intestinal microbiota with regards to RA treatment has been demonstrated: in RA patients who are undergoing treatment with DMARDS, the gut microbiota is more similar to healthy controls as opposed to other as yet untreated RA

patients (Zhang et al. 2015), although this has not yet been confirmed by longitudinal studies (discussed in Chapter 2, Chapter 5 and Chapter 6). *Bacteroides* is a dominant taxon which has been demonstrated to promote an anti-inflammatory milieu via induction of T-regulatory cells. Potentially, as posited in Zhang *et al.,* some of the anti-inflammatory effects of DMARDs may act via the gut microbiota. The results presented in Chapter 6 would support this hypothesis with regards to *Bacteroides.* The other taxa associations require further functional work to provide insight into pathological relevance.

8.2.4 Specific aims: Aim 4

4) Investigate the oral microbiota association with RA with regards to host genetic factors.

In Chapter 7, I set out to address the fourth aim of this thesis. To achieve this, I firstly characterised the oral microbiota with regards to potentially influential biological and technical factors, and secondly I modelled the association of the oral microbiota with genetic risk of RA. In undertaking the first task, I demonstrated that the saliva microbiota is substantially heritable. 113 ASVs were significantly heritable. The most heritable phylum was Firmicutes, and the three most heritable taxa were *Veillonella atypica*, *Campylobacter concisus*, and *Selenomonas_3.* Diversity of the oral microbiota was associated with age and fasting prior to sample collection. Conversely - and interestingly - there was no association with diet, frailty, BMI or any of the other covariates considered. The oral microbiota is little understood compared to the gut microbiota, and studies which elucidate the modulatory factors are valuable at this stage. The results of the Chapter 7 are therefore informative for the wider research area. To my knowledge, the association of age with oral microbiota diversity, even after controlling for other factors including general health is novel.

With regards to the association of genetic risk of RA with the oral microbiota, a taxonomic association was shown. However, the relevance of these taxa to the pathogenesis of RA

was not supported or refuted by the prior literature. Overall, I suggest that the study is presently not convincing with regards to a potential influence of host genetic factors on the oral microbiota in RA (discussed in Chapter 7).

8.3 Consideration of results within the context of RA pathogenesis

An overarching question in RA microbiome research is whether the microbiota contribute to RA pathogenesis. Specifically, whether the microbiota could constitute a causative immunomodulatory influence, contributing to the initial development of autoimmunity which may begin decades before the development of clinical RA. This is a complex question to resolve, which would potentially require interventions over many years, being logistically and ethically difficult to conduct. Therefore, the research field needed evidence of associations at different stages in the development of RA to focus both mechanistic experiments and further detailed observational studies in the next stage.

Herein lies the key strength of the epidemiological approach of this thesis. By looking at those with genetic risk but no RA, I was able to interrogate changes which could appear before the clinical diagnosis; by looking at a preclinical RA cohort with increased genetic risk (first degree relatives), we were able to clarify changes at onset of symptoms; finally I was able to look at changes in established RA, by comparing the gut microbiota of affected twins to their identical twin sisters. Thus, the thesis has added key observational data which enable clarifications to the fields summarized in Figure 8.1. Clearly, as observational evidence, this does not provide evidence of causation of any changes seen at any point along the time-course. Further studies are needed to establish this.

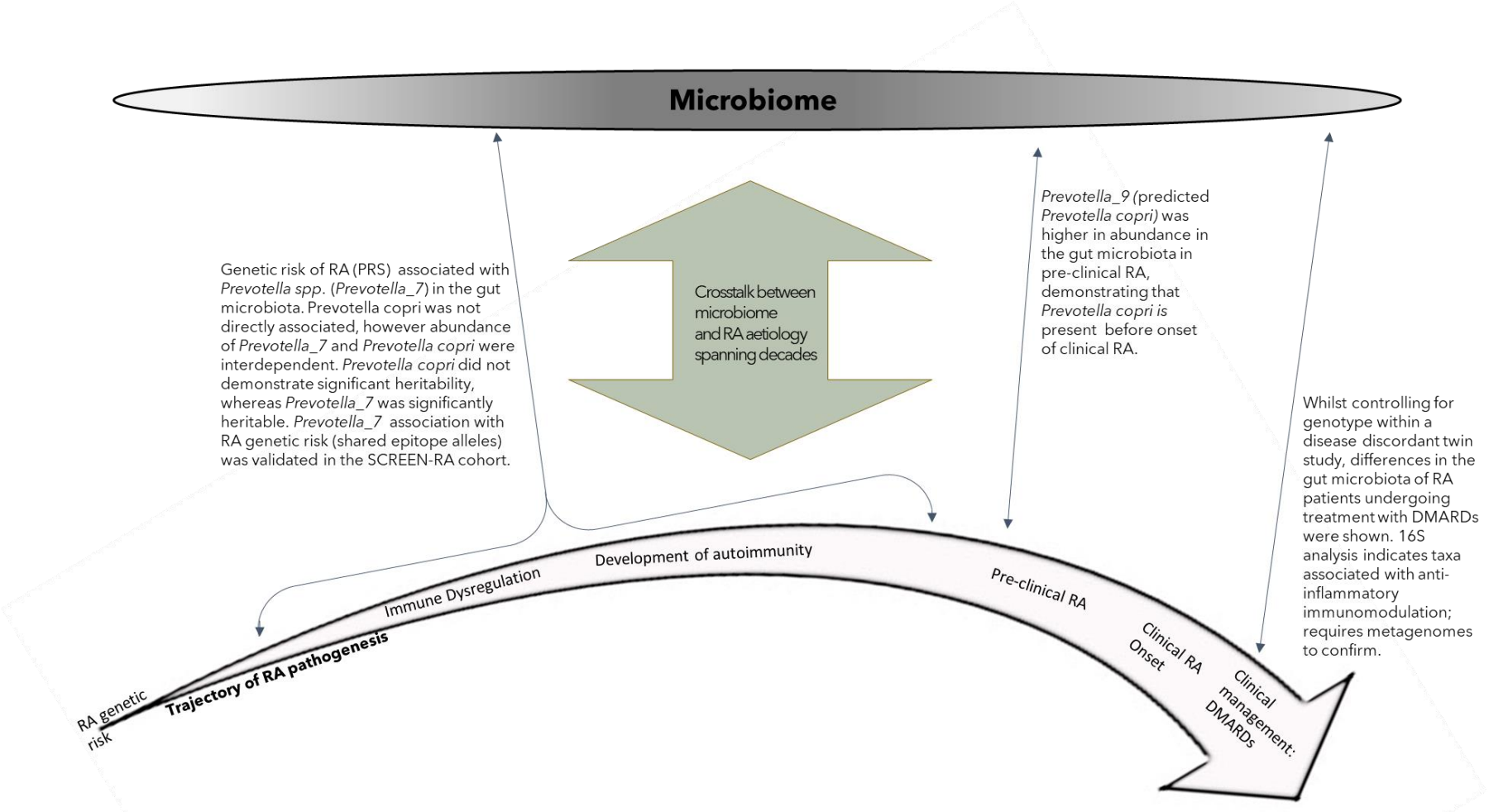


Figure 8.1 Summary of clarification to the research area provided by this work

The work in Chapter 5 and Chapter 7 relating genetic risk of RA to the gut (stool) and oral (saliva) microbiota, respectively, demonstrated a link between genetic risk without the actual disease and taxa which have been previously associated with RA . My study focused on adults in the 55-70 age range. As host genotype is stable across the life-course, and therefore such genotype-microbiota associations may hold true even from late childhood, well before any onset of clinical disease in those genetically at risk. The work presented in this thesis is the first to provide evidence that pathophysiologically pertinent differences in the microbiota may potentially occur prior to the development of RA symptoms placing microbiota changes plausibly early in the course of RA development. The evidence was stronger for the gut microbiome; the association demonstrated with *Prevotella spp.* in particular will be of interest to the wider RA microbiome research community (discussed in detail in Chapter 5). Whilst further work is needed, this study bolsters the feasibility of the microbiome as a target in pre-clinical RA, for prevention of progression to clinical RA of arthralgia patients (Cope 2017).

The study of the microbiota of RA discordant twins, presented in Chapter 6 is the largest study to date of the microbiome of established RA patients. This work suggests that genetic and dietary factors are superseded by xenobiotic factors in the establishment of gut microbiota composition in RA patients. The study design uniquely afforded for control of host genetic factors. However, it is difficult to extrapolate findings to the earlier trajectory of RA development due to the effects of RA treatment on the microbiota; participants of the study were established RA patients undergoing treatment with DMARDs. The contrast between taxon differences observed in early (untreated RA) in prior studies, and in established RA in the present study highlight that DMARDs are likely to be a driving factor of microbiome composition in RA patients. It is clear that there is a need to investigate the microbiota of RA patients from the early stages before commencement of treatment, following through to longitudinally assess microbiome changes after DMARDs; I will be investigating this as part of Frances Williams' team following completion of my PhD. The evidence base for this approach is further described in Section 8.5.

8.4 Limitations of methods

With regards to the work presented in this thesis, I must first acknowledge that just one chapter, Chapter 6, included investigation of participants who were diagnosed with RA. Whilst this was appropriate for the research questions addressed in this thesis, use of additional datasets of RA cases would increase the scope of this work. Limitations of the methods for each study are discussed in the respective chapters. Here I include a general discussion of the limits of 16S methods which are used throughout the thesis.

8.4.1 Limitations of 16S rRNA gene sequencing

16S rRNA gene sequencing is the most frequently used method for characterising the microbiome. The main alternate approach of metagenome sequencing yields more information, but is inhibited by economic factors, particularly in application to larger sets of samples. 16S methods provide a balance in terms of information and economics. The generation of ASVs from 16S yields useful and valid information regarding the taxa present at genus and to a lesser extent species level – the latter is now much improved following the introduction of ASV methods. However, there are important limitations to 16S data which are described below. Some of these limitations are common to high throughput sequencing, including metagenomic methods.

The most cited limitation of 16S data is that it relies on a single marker gene, and secondary to this the method does not usually allow for taxonomic identification at the species or strain level. Further, and perhaps more frustratingly, the specific genes and enzymes present for each bacterium cannot be identified - these can vary substantially within members of the same taxon. As such, using 16S data, it is difficult to answer the question of 'which bacteria are present?', and largely impossible to answer the question of 'what are they doing?'.

However, the limitations of 16S data go further than this due to: variability in constant regions; 16S rRNA gene copy number variations; lack of variability in 'variable region' of some species; reliance on databases; chimera formation and difficulty in effective statistical management of these artefacts; abundance estimation inaccuracy; sequencing error. Collectively, these account for the lower reliability of 16S data compared to metagenomics data.

The first issue is that the constant regions are captured with variable success by the degenerate primers which are utilised. Degenerate primers are defined as a mix of oligonucleotide sequences in which some positions contain a number of possible bases, therefore comprising a population of primers with similar sequences which cover all possible nucleotide combinations for a peptide sequence (Li, Hingamp, et al. 2018). The taxa detected by 16S data is dependent on the primers which are chosen. In addition, depending on the variable region targeted, primers can be ineffective for binding the variable region of some taxa. Indeed, the V4 variable region completely fails to amplify *Cutibacterium acnes* (Long et al. 2019).

Secondly, 16S relies on detection of 16S rRNA genes, however there can be multiple of these genes within the same genome – this is referred to as copy number. The multiplicity is not constant across taxa, and therefore this is a source of bias. It has been demonstrated members of Firmicutes and Gammaproteobacteria can possess up to 6 copies of the 16S rRNA gene, whilst members of Bacteroidetes possess up to 3 copies (Sun et al. 2013).

Third, the variable region is not variable enough to distinguish some species. *Escherechia spp*. and *Shigella spp*. have distinct natural histories, but it is impossible however to separate these taxa based on their 16S genes - all of their variable regions are too similar (Devanga Ragupathi et al. 2017).

Fourth, taxonomic identification is reliant on databases. These resources are improving, however we still have not characterised at least 50% of existing taxa, and of those that are

classified, some are incorrectly or incompletely labelled in some instances (Konstantinidis et al. 2017). Researchers are working hard to curate these invaluable resources and advancements are being made. In this thesis, the advancements in resolution of taxon annotations in the SILVA database and research relevance of these have been demonstrated.

Fifth, an unavoidable pitfall of the method is chimera formation. Chimeras are artefacts of the PCR process - artificial 16S sequences which are derived from multiple sources. Chimera formation may be caused by incomplete primer extension, disassociation of polymerase, and misincorporation of polymerase and secondary structures. The inclusion of these in the final dataset can produce spurious results, such as artificial diversity, incorrect taxon identification and skewed phylogenetic trees (Parks et al. 2018). 16S data processing methods for OTUs and ASVs include steps to exclude chimeras, however these are not completely effective (Callahan et al. 2016).

Chimera exclusion steps are typically based on sequence comparison or low abundance. However database sequences can be inaccurate, and chimeras which were formed early in the PCR process may not be low abundance. In addition, the issues described above pertaining to PCR, choice of primers and copy number of the 16S rRNA gene confer that abundance estimates in 16S data are fallible. The abundance estimates are skewed, and therefore should be interpreted with caution.

It is worth noting that metagenomics methods also hold limitations, however these are also common to 16S data. Many of these issues are therefore unavoidable using current methods. However key advantages of metagenomic methods when compared to 16S is detection of bacterial genes present, giving invaluable insight into functionality - likely to be more relevant than specific taxa - and metagenomics methods allow far greater detection bandwidth beyond the genus level.

8.5 Future directions

A discussion of the development of the specific studies in this thesis are included in the respective chapters. Here, I suggest how the research area as a whole could be advanced. As the research field is in the exploratory phase, primarily, additional epidemiological studies, in are required in order to characterise the microbiota in RA. In order to understand pathological relevance of microbiota differences associated with RA, we must develop a robust understanding of the nature of these differences; as discussed in Chapter 4 and Chapter 6, there are contrasting findings in prior studies. Replication of the current findings in other cohorts will be important. Such studies should include consideration where possible of confounding factors.

In terms of furthering understanding of the trajectory of the link between the microbiome and RA, it will be important to consider temporal differences in the microbiota occurring during the life-course. Development of RA occurs over decades, and aetiological aggravation via the intestinal microbiota may be evident some time before onset of clinical symptoms. To investigate this, the most informative approach would a prospective longitudinal study of participants at higher risk of RA, for example the children of RA patients. However this would be an extensive and long term study. There are two alternative approaches with less logistical barriers. The first is investigation of the microbiome as part of large studies such as UK biobank, in which individuals are already being prospectively followed. The second is to take a similar approach as presented in Chapter 5 and Chapter 7, to assess association of genetic risk of RA with the commensal microbiota. For example, it would be interesting to investigate association of genetic risk of RA and the microbiota in children from the ALSPAC cohort. Here, it would interesting to see if the *Prevotella* association replicated in these younger participants. This would give impetus to the theory that an increase in *Prevotella spp*. lies earlier on the in the trajectory of RA pathogenesis and is more likely to be causal.

In understanding the trajectory of microbiome involvement in RA, it will also be important to consider other non-genetic risk factors for RA. With regard to the oral microbiome, a first step would be to investigate concurrent periodontal disease with RA in relation to oral microbiome and potential for shared genetic aetiology with RA, and the microbiome in relation to RA treatment response. With regards to the second suggestion, further studies are required of the oral microbiota association with periodontal disease. In order to demonstrate potential overlap in genetic aetiology, bivariate classical twin modelling would be informative.

The link between the microbiome and RA extends beyond the microbiota differences which potentially contribute to the pathophysiology of RA; an important area which was outside the scope of this thesis, but which should be addressed in future studies is the implication of the microbiota in efficacy of treatments. There is substantial variation between patients in the response to treatment, and this may be partially mediated via the microbiota (Sayers et al. 2018). With regards to Methotrexate response, only 31% patients respond satisfactorily after 3 months according to the EULAR criteria (Lie et al. 2012), and attempts to reliably stratify response based on age, sex, BMI, smoking history, genetic factors, and serum biomarkers have proved unsuccessful (Wijbrandts and Tak 2017, Sayers et al. 2018).

There are several examples of mechanisms by which the microbiota influence drug bioavailability, efficacy, or toxicity. In instances where bioavailability is dependent on internal metabolism, drugs are referred to as prodrugs. Sulfasalazine, a common RA therapeutic, is a prodrug which is functionally dependent on the intestinal microbiota. It is converted to its bioavailable form via gut microbiota azo reductase (Sayers et al. 2018). Conversely, as discussed previously, medications may alter the composition of the microbiota. Methotrexate, which is the first line treatment in RA, exhibits anti-microbial properties. The mechanism of action of methotrexate is folate inhibition, however the pathways involved are not specific to human cells, and also affect the microbiota. In RA this is suggested to be a potential mechanism of action, as RA therapeutics are associated with a

microbiota more similar to non-diseased than diseased individuals (Zhang et al. 2015). Work is underway to understand the longitudinal response to RA treatment in relation to the microbiome.

When studying the microbiome, computational and wet-lab based approaches go hand in hand; computational data is generated using wet-lab techniques. However, there are numerous primarily wet lab approaches which will be required alongside computational methods to address the complex task of investigating the functional implication of the microbiota in RA pathogenesis. Studies of immune reactivity to RA associated taxa in RA patients and control participants will be informative. I am aware of just one such study to date which showed promising results: IgA or IgG immune reactivity for *P.copri* was demonstrated in 24% of RA patients compared to 2% of healthy control participants (Pianta et al. 2016). Further studies using a similar approach would be informative. Well-designed studies using animal models such as murine models can be valuable, especially when investigating a specific aspect of microbiota-host physiology interaction. These approaches will need to be applied to potentially RA relevant microbiota identified through computational methods. This will include *Prevotella spp* however the taxa implicated and their microbial community networks are likely to be much broader than these candidate taxa.. A more novel approach is the use of organoids to model functionality relevant to humans in vitro. There is also gathering interest in bacterial constituents such as OMVs which wet lab methods will be required in order to understand (Jones et al. 2020).

As discussed in Chapter 1, OMVs which are produced ubiquitously by gram negative bacteria have been demonstrated to modify the local environment to facilitate bacterial proliferation, signal between bacterial species and have in addition been shown to communicate directly with host cells (Ellis and Kuehn 2010, Kulkarni and Jagannadham 2014, Stentz et al. 2014). In the host, OMVs affect intracellular signalling and overall metabolic profile (Stentz et al. 2014, Bryant et al. 2017a). Investigation of OMVs may be an important step in understanding the link between the microbiome and host. In support of this, outer membrane vesicles of pathogenic and non-pathogenic strains of the same

species have shown differing metabolic associations (Zakharzhevskaya et al. 2017a). *P.copri* is a gram-negative bacterium which produces OMVs, however the potential role of *P.copri* derived OMVs in relation to host inflammation and immunity has not yet been reported. OMVs of *P.copri* isolated from RA patients may have differences in content from other RA associated taxa, may exacerbate the pro-inflammatory host environment in RA. It would be informative to explore the content of outer membrane vesicles of *P.copri* using high performance liquid chromatography (HPLC). In *P.copri* strains derived from new-onset RA patients, HPLC could be used to compare P.copri OMVs to OMVs of P.copri from RA unaffected people who also have a high abundance of this taxon. Similarly, it would be useful to compare to OMVs from other bacteria, to give insight into differential effects on the host and potential role in inflammation.

With regards to microbiome studies in general, where possible studies should use metagenomes methods, in order to improve detection of taxa and specific genes present in samples. Demonstrating the specific genes present is important, as there is a substantially higher degree of shared functionality between individuals than shared taxa. Studies should also include concurrent measures which can demonstrate functionality of the microbiota: metabolomics and meta-transcriptomics. In addition, it will be valuable to consider non-bacterial constituents of the microbiome, such as fungi and viruses. Viruses - including bacteriophages - are the most abundant constituent of the microbiome. Methods for investigating the virome and particular the role of bacteriophages are not well established as yet but will be important moving forward.

8.6 Conclusions

The work presented in this thesis constitutes several novel findings which collectively support the hypothesis that the commensal microbiota have a role in the aetiology of RA. In particular, RA characteristic microbiota differences are indicated to occur prior to onset of clinical RA, and associate with host genetic factors. This work provides support for the

concept of the microbiome as a feasible RA-preventative target for patients with arthralgia. Prospectively, this work paves the way for further study of microbiota implication across the trajectory of RA pathogenesis spanning from earlier stages of immune dysregulation, to response to RA therapeutics in those with established disease. Further, this work provided several findings of wider interest beyond the scope of RA microbiome, which provide insight into influential factors of the oral and gut microbiota.

References

Albani, S., Keystone, E. C., Nelson, J. L., Ollier, W. E., La Cava, A., Montemayor, A. C., Weber, D. A., Montecucco, C., Martini, A., and Carson, D. A., 1995. Positive selection in autoimmunity: abnormal immune responses to a bacterial dnaJ antigenic determinant in patients with early rheumatoid arthritis. *Nature Medicine*, 1 (5), 448–452.

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D., and Finn, R. D., 2019. A new genomic blueprint of the human gut microbiota. *Nature*, 568 (7753), 499–504.

Alpizar-Rodriguez, D., Lesker, T. R., Gronow, A., Gilbert, B., Raemy, E., Lamacchia, C., Gabay, C., Finckh, A., and Strowig, T., 2019. Prevotella copri in individuals at risk for rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 78 (5), 590–593.

Alseth, E. O., Pursey, E., Luján, A. M., McLeod, I., Rollie, C., and Westra, E. R., 2019. Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature*, 574 (7779), 549–552.

Anderson, M. J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26 (1), 32–46.

Badri, M., Kurtz, Z. D., Bonneau, R., and Müller, C. L., 2020. Shrinkage improves estimation of microbial associations under different normalization methods. *bioRxiv*, 406264.

Bates, D., Mächler, M., Bolker, B., and Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67 (1), 1–48.

Bauché, D. and Marie, J. C., 2017. Transforming growth factor β: a master regulator of the gut microbiota and immune cell interactions. *Clinical & Translational Immunology*, 6 (4), e136.

Belkaid, Y. and Hand, T., 2014. Role of the Microbiota in Immunity and inflammation. *Cell*, 157 (1), 121–141.

Belkaid, Y. and Harrison, O. J., 2017. Homeostatic Immunity and the Microbiota. *Immunity*, 46 (4), 562–576.

Bernard, N. J., 2018. Synergy of environmental and genetic risk in RA. *Nature Reviews Rheumatology*, 14 (6), 319.

Bingham, C. O. and Moni, M., 2013. Periodontal disease and rheumatoid arthritis: the evidence accumulates for complex pathobiologic interactions. *Current opinion in rheumatology*, 25 (3), 345–353.

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., Spector, T. D., Keinan, A., Ley, R. E., Gevers, D., and Clark, A. G., 2015. Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, 16 (1), 191.

Bluett, J., Owen, S.-A., Massey, J., Alfirevic, A., Pirmohamed, M., Plant, D., Verstappen, S. M. M., Barton, A., and Pneumonitis Study Consortium, 2017. HLA-A 31:01 is not associated with the development of methotrexate pneumonitis in the UK population: results from a genome-wide association study. *Annals of the Rheumatic Diseases*, 76 (12), e51.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E., Silva, R. D., Dorrestein, P. C., Douglas, G. M., Durall, D. M., Duvallet, C., Edwardson, C. F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J. M., Gibson, D. L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G., Janssen, S., Jarmusch, A. K., Jiang, L., Kaehler, B., Kang, K. B., Keefe, C. R., Keim, P., Kelley, S. T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M. G., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B. D., McDonald, D., McIver, L. J., Melnik, A. V., Metcalf, J. L., Morgan, S. C., Morton, J., Naimey, A. T., Navas-Molina, J. A., Nothias, L. F., Orchanian, S. B., Pearson, T., Peoples, S. L., Petras, D., Preuss, M. L., Pruesse, E., Rasmussen, L. B., Rivers, A., Michael S Robeson, I. I., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S. J., Spear, J. R., Swafford, A. D., Thompson, L. R., Torres, P. J., Trinh, P., Tripathi, A., Turnbaugh, P. J., Ul-Hasan, S., Hooft, J. J. van der, Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., Hippel, M. von, Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K. C., Williamson, C. H., Willis, A. D., Xu, Z. Z., Zaneveld, J. R., Zhang, Y., Zhu, Q., Knight, R., and Caporaso, J. G., 2018. *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science* [online]. PeerJ Inc. No. e27295v2. Available from: https://peerj.com/preprints/27295 [Accessed 4 Jul 2020].

Bossini-Castillo, L., de Kovel, C., Kallberg, H., van 't Slot, R., Italiaander, A., Coenen, M., Tak, P. P., Posthumus, M. D., Wijmenga, C., Huizinga, T., van der Helm-van Mil, A. H. M., Stoeken-Rijsbergen, G., Rodriguez-Rodriguez, L., Balsa, A., González-Álvaro, I., González-Gay, M. Á., Gómez-Vaquero, C., Franke, B., LifeLines Cohort Study, Vermeulen, S., van der Horst-Bruinsma, I., Dijkmans, B. a. C., Wolbink, G. J., Ophoff, R. A., Maehlen, M. T., van Riel, P., Merriman, M., Klareskog, L., Lie, B. A., Merriman, T., Crusius, J. B. A., Brouwer, E., Martin, J., de Vries, N., Toes, R., Padyukov, L., and Koeleman, B. P. C., 2015. A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides. *Annals of the Rheumatic Diseases*, 74 (3), e15.

Bowyer, R. C. E., Jackson, M. A., Pallister, T., Skinner, J., Spector, T. D., Welch, A. A., and Steves, C. J., 2018. Use of dietary indices to control for diet in human gut microbiota studies. *Microbiome*, 6 (1), 77.

Bray, J. R. and Curtis, J. T., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27 (4), 325–349.

Broussard, J. L. and Devkota, S., 2016. The changing microbial landscape of Western society: Diet, dwellings and discordance. *Molecular Metabolism*, 5 (9), 737–742.

Bryant, W. A., Stentz, R., Le Gall, G., Sternberg, M. J. E., Carding, S. R., and Wilhelm, T., 2017a. In Silico Analysis of the Small Molecule Content of Outer Membrane Vesicles Produced by Bacteroides thetaiotaomicron Indicates an Extensive Metabolic Link between Microbe and Host. *Frontiers in Microbiology* [online], 8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02440/full [Accessed 21 May 2018].

Bryant, W. A., Stentz, R., Le Gall, G., Sternberg, M. J. E., Carding, S. R., and Wilhelm, T., 2017b. In Silico Analysis of the Small Molecule Content of Outer Membrane Vesicles Produced by Bacteroides thetaiotaomicron Indicates an Extensive Metabolic Link between Microbe and Host. *Frontiers in Microbiology* [online], 8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02440/full [Accessed 5 Jun 2018].

Bullock, J., Rizvi, S. A. A., Saleh, A. M., Ahmed, S. S., Do, D. P., Ansari, R. A., and Ahmed, J., 2019. Rheumatoid Arthritis: A Brief Overview of the Treatment. *Medical Principles and Practice*, 27 (6), 501–507.

Bunyavanich, S., Silberg, J. L., Lasky-Su, J., Gillespie, N. A., Lange, N. E., Canino, G., and Celed**ó**n, J. C., 2013. A Twin Study of Early-Childhood Asthma in Puerto Ricans. *PLOS ONE*, 8 (7), e68473.

Burn, G. L., Svensson, L., Sanchez-Blanco, C., Saini, M., and Cope, A. P., 2011. Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? *FEBS Letters*, 585 (23), 3689–3698.

Callahan, B., 2018. Silva taxonomic training data formatted for DADA2 (Silva version 132). [online]. Available from: https://zenodo.org/record/1172783#.WyeylanTWHs [Accessed 18 Jun 2018].

Callahan, B. J., McMurdie, P. J., and Holmes, S. P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11 (12), 2639–2643.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P., 2016. DADA2: High resolution sample inference from Illumina amplicon data. *Nature methods*, 13 (7), 581–583.

Cao, Y., Fanning, S., Proos, S., Jordan, K., and Srikumar, S., 2017. A Review on the Applications of Next Generation Sequencing Technologies as Applied to Food-Related Microbiome Studies. *Frontiers in Microbiology* [online], 8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01829/full [Accessed 19 May 2020].

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108 (Supplement 1), 4516–4522.

Carmody, R. N. and Turnbaugh, P. J., 2014. Host-microbial interactions in the metabolism of therapeutic and diet-derived xenobiotics. *The Journal of Clinical Investigation*, 124 (10), 4173–4181.

Ceccarelli, F., Orrù, G., Pilloni, A., Bartosiewicz, I., Perricone, C., Martino, E., Lucchetti, R., Fais, S., Vomero, M., Olivieri, M., di Franco, M., Priori, R., Riccieri, V., Scrivo, R., Shoenfeld, Y., Alessandri, C., Conti, F., Polimeni, A., and Valesini, G., 2018. Porphyromonas gingivalis in the tongue biofilm is associated with clinical outcome in rheumatoid arthritis patients. *Clinical and Experimental Immunology*, 194 (2), 244–252.

Cecchi, I., Arias de la Rosa, I., Menegatti, E., Roccatello, D., Collantes-Estevez, E., Lopez-Pedrera, C., and Barbarroja, N., 2018. Neutrophils: Novel key players in Rheumatoid Arthritis. Current and future therapeutic targets. *Autoimmunity Reviews* [online]. Available from: http://www.sciencedirect.com/science/article/pii/S1568997218302167 [Accessed 29 Sep 2018].

Chalan, P., van den Berg, A., Kroesen, B.-J., Brouwer, L., and Boots, A., 2015. Rheumatoid Arthritis, Immunosenescence and the Hallmarks of Aging. *Current Aging Science*, 8 (2), 131–146.

Chao, A., 1984. Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, 11 (4), 265–270.

Cheignon, C., Tomas, M., Bonnefont-Rousselot, D., Faller, P., Hureau, C., and Collin, F., 2018. Oxidative stress and the amyloid beta peptide in Alzheimer's disease. *Redox Biology*, 14, 450–464.

Chen, B., Zhao, Y., Li, S., Yang, L., Wang, H., Wang, T., Bin Shi, Gai, Z., Heng, X., Zhang, C., Yang, J., and Zhang, L., 2018. Variations in oral microbiome profiles in rheumatoid arthritis and osteoarthritis with potential biomarkers for arthritis screening. *Scientific Reports* [online], 8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6244360/ [Accessed 4 Sep 2019].

Chen, C.-K., Wu, Y.-T., and Chang, Y.-C., 2017. Association between chronic periodontitis and the risk of Alzheimer's disease: a retrospective, population-based, matched-cohort study. *Alzheimer's Research & Therapy* [online], 9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5547465/ [Accessed 19 Jun 2020].

Chen, T., Yu, W.-H., Izard, J., Baranova, O. V., Lakshmanan, A., and Dewhirst, F. E., 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* [online], 2010. Available from: https://academic.oup.com/database/article/doi/10.1093/database/baq013/405450 [Accessed 20 Jun 2020].

Chial, H., 2008. *DNA Sequencing Technologies Key to the Human Genome Project* [online]. Human Genome Project: Sequencing the Human Genome | Learn Science at Scitable. Available from: https://www.nature.com/scitable/topicpage/dna-sequencing-technologies-key-to-the-human-828/ [Accessed 24 Jun 2020].

Colombo, V., Magalhães, C. B., Hartenbach, F. A. R. R., Martins do Souto, R., and Maciel da Silva-Boghossian, C., 2016. Periodontal-disease-associated biofilm: A reservoir for pathogens of medical importance. *Microbial Pathogenesis*, 94, 27–34.

Combellick, J. L., Shin, H., Shin, D., Cai, Y., Hagan, H., Lacher, C., Lin, D. L., McCauley, K., Lynch, S. V., and Dominguez-Bello, M. G., 2018. Differences in the fecal microbiota of neonates born at home or in the hospital. *Scientific Reports*, 8 (1), 1–9.

Condic, M., Oberstein, T. J., Herrmann, M., Reimann, M. C., Kornhuber, J., Maler, J. M., and Spitzer, P., 2014. N-truncation and pyroglutaminylation enhances the opsonizing capacity of Aβ-peptides and facilitates phagocytosis by macrophages and microglia. *Brain, Behavior, and Immunity*, 41, 116–125.

Cope, A. P., 2017. Emerging therapies for pre-RA. *Best Practice & Research Clinical Rheumatology* [online]. Available from: http://www.sciencedirect.com/science/article/pii/S1521694217300153.

Cui, J., Stahl, E. A., Saevarsdottir, S., Miceli, C., Diogo, D., Trynka, G., Raj, T., Mirkov, M. U., Canhao, H., Ikari, K., Terao, C., Okada, Y., Wedrén, S., Askling, J., Yamanaka, H., Momohara, S., Taniguchi, A., Ohmura, K., Matsuda, F., Mimori, T., Gupta, N., Kuchroo, M., Morgan, A. W., Isaacs, J. D., Wilson, A. G., Hyrich, K. L., Herenius, M., Doorenspleet, M. E., Tak, P.-P., Crusius, J. B. A., van der Horst-Bruinsma, I. E., Wolbink, G. J., van Riel, P. L. C. M., van de Laar, M., Guchelaar, H.-J., Shadick, N. A., Allaart, C. F., Huizinga, T. W. J., Toes, R. E. M., Kimberly, R. P., Bridges, S. L., Criswell, L. A., Moreland, L. W., Fonseca, J. E., de Vries, N., Stranger, B. E., De Jager, P. L., Raychaudhuri, S., Weinblatt, M. E., Gregersen, P. K., Mariette, X., Barton, A., Padyukov, L., Coenen, M. J. H., Karlson, E. W., and Plenge, R. M., 2013. Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. *PLoS genetics*, 9 (3), e1003394.

Cui, J., Taylor, K. E., Destefano, A. L., Criswell, L. A., Izmailova, E. S., Parker, A., Roubenoff, R., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Karlson, E. W., 2009. Genome-wide association study of determinants of anti-cyclic citrullinated peptide antibody titer in adults with rheumatoid arthritis. *Molecular Medicine (Cambridge, Mass.)*, 15 (5–6), 136–143.

Dani, S., Prabhu, A., Chaitra, K. R., Desai, N. C., Patil, S. R., and Rajeev, R., 2016. Assessment of Streptococcus mutans in healthy versus gingivitis and chronic periodontitis: A clinico-microbiological study. *Contemporary Clinical Dentistry*, 7 (4), 529–534.

Dao, M. C., Everard, A., Aron-Wisnewsky, J., Sokolovska, N., Prifti, E., Verger, E. O., Kayser, B. D., Levenez, F., Chilloux, J., Hoyles, L., Consortium, M.-O., Dumas, M.-E., Rizkalla, S. W., Doré, J., Cani, P. D., and Clément, K., 2015. Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut*, gutjnl-2014-308778.

Davenport, E. R., 2017. Tooth Be Told, Genetics Influences Oral Microbiome. *Cell Host & Microbe*, 22 (3), 251–253.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J.,

and Turnbaugh, P. J., 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505 (7484), 559–563.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., and Callahan, B. J., 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, 6 (1), 226.

De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N., and Ercolini, D., 2019. Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets. *Cell Host & Microbe*, 25 (3), 444-453.e3.

Devanga Ragupathi, N. K., Muthuirulandi Sethuvel, D. P., Inbanathan, F. Y., and Veeraraghavan, B., 2017. Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. *New Microbes and New Infections*, 21, 58–62.

Dominy, S. S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., Nguyen, M., Haditsch, U., Raha, D., Griffin, C., Holsinger, L. J., Arastu-Kapur, S., Kaba, S., Lee, A., Ryder, M. I., Potempa, B., Mydel, P., Hellvard, A., Adamowicz, K., Hasturk, H., Walker, G. D., Reynolds, E. C., Faull, R. L. M., Curtis, M. A., Dragunow, M., and Potempa, J., 2019. Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Science Advances*, 5 (1), eaau3333.

Donia, M. S. and Fischbach, M. A., 2015. HUMAN MICROBIOTA. Small molecules from the human microbiota. *Science (New York, N.Y.)*, 349 (6246), 1254766.

Dorrestein, P. C., Mazmanian, S. K., and Knight, R., 2014. Finding the missing links among metabolites, microbes, and the host. *Immunity*, 40 (6), 824–832.

Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., and Alm, E. J., 2017. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature Communications*, 8 (1), 1784.

Edgar, R. C., 2017. Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ*, 5, e3889.

Eerola, E., Möttönen, T., Hannonen, P., Luukkainen, R., Kantola, I., Vuori, K., Tuominen, J., and Toivanen, P., 1994. Intestinal flora in early rheumatoid arthritis. *British Journal of Rheumatology*, 33 (11), 1030–1038.

Ellis, T. N. and Kuehn, M. J., 2010. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. *Microbiology and molecular biology reviews: MMBR*, 74 (1), 81–94.

Emery, D. C., Shoemark, D. K., Batstone, T. E., Waterfall, C. M., Coghill, J. A., Cerajewska, T. L., Davies, M., West, N. X., and Allen, S. J., 2017. 16S rRNA Next Generation Sequencing Analysis Shows Bacteria in Alzheimer's Post-Mortem Brain. *Frontiers in Aging Neuroscience* [online], 9. Available from: https://www.frontiersin.org/articles/10.3389/fnagi.2017.00195/full [Accessed 6 Feb 2019].

Eriksson, K., Fei, G., Lundmark, A., Benchimol, D., Lee, L., Hu, Y. O. O., Kats, A., Saevarsdottir, S., Catrina, A. I., Klinge, B., Andersson, A. F., Klareskog, L., Lundberg, K., Jansson, L., and Yucel-Lindberg, T., 2019. Periodontal Health and Oral

Microbiota in Patients with Rheumatoid Arthritis. *Journal of Clinical Medicine*, 8 (5), 630.

Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K., Amos, C. I., Padyukov, L., Toes, R. E. M., Huizinga, T. W. J., Wijmenga, C., Trynka, G., Franke, L., Westra, H.-J., Alfredsson, L., Hu, X., Sandor, C., de Bakker, P. I. W., Davila, S., Khor, C. C., Heng, K. K., Andrews, R., Edkins, S., Hunt, S. E., Langford, C., Symmons, D., Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate, Wellcome Trust Case Control Consortium, Concannon, P., Onengut-Gumuscu, S., Rich, S. S., Deloukas, P., Gonzalez-Gay, M. A., Rodriguez-Rodriguez, L., Ärlsetig, L., Martin, J., Rantapää-Dahlqvist, S., Plenge, R. M., Raychaudhuri, S., Klareskog, L., Gregersen, P. K., and Worthington, J., 2012. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genetics*, 44 (12), 1336–1340.

Faith, D. P., 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61 (1), 1–10.

Fernández, M., Gobartt, A. L., Balañá, M., and COOPERA Study Group, 2010. Behavioural symptoms in patients with Alzheimer's disease and their association with cognitive impairment. *BMC neurology*, 10, 87.

Fierer, N., Ferrenberg, S., Flores, G. E., González, A., Kueneman, J., Legg, T., Lynch, R. C., McDonald, D., Mihaljevic, J. R., O'Neill, S. P., Rhodes, M. E., Song, S. J., and Walters, W. A., 2012. From Animalcules to an Ecosystem: Application of Ecological Concepts to the Human Microbiome. *Annual Review of Ecology, Evolution, and Systematics*, 43 (1), 137–155.

Finckh, A., Courvoisier, D., and Lamacchia, C., 2020. Measuring ACPA in the general population or primary care: is it useful? *RMD Open*, 6 (1), e001085.

Firestein, G. S. and McInnes, I. B., 2017. Immunopathogenesis of Rheumatoid Arthritis. *Immunity*, 46 (2), 183–196.

Foley, M. H., O'Flaherty, S., Barrangou, R., and Theriot, C. M., 2019. Bile salt hydrolases: Gatekeepers of bile acid metabolism and host-microbiome crosstalk in the gastrointestinal tract. *PLoS Pathogens* [online], 15 (3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6405046/ [Accessed 8 Apr 2020].

Freudenberg, J., Lee, H.-S., Han, B.-G., Shin, H. D., Kang, Y. M., Sung, Y.-K., Shim, S.-C., Choi, C.-B., Lee, A. T., Gregersen, P. K., and Bae, S.-C., 2011. Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis and Rheumatism*, 63 (4), 884–893.

Friedman, J., Higgins, L. M., and Gore, J., 2017. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution*, 1, 0109.

Gao, L., Xu, T., Huang, G., Jiang, S., Gu, Y., and Chen, F., 2018. Oral microbiomes: more and more importance in oral cavity and whole body. *Protein & Cell*, 9 (5), 488–500.

Gavina, M. K. A., Tahara, T., Tainaka, K., Ito, H., Morita, S., Ichinose, G., Okabe, T., Togashi, T., Nagatani, T., and Yoshimura, J., 2018. Multi-species coexistence in Lotka-Volterra competitive systems with crowding effects. *Scientific Reports*, 8 (1), 1198.

Gibofsky, A., 2014. Epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis: A Synopsis. *The American Journal of Managed Care*, 20 (7 Suppl), S128-135.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* [online], 8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02224/full#F2 [Accessed 30 Apr 2020].

Gomez, A., Espinoza, J. L., Harkins, D. M., Leong, P., Saffery, R., Bockmann, M., Torralba, M., Kuelbs, C., Kodukula, R., Inman, J., Hughes, T., Craig, J. M., Highlander, S. K., Jones, M. B., Dupont, C. L., and Nelson, K. E., 2017. Host Genetic Control of the Oral Microbiome in Health and Disease. *Cell Host & Microbe*, 22 (3), 269-278.e3.

Gomez, A., Luckey, D., Yeoman, C. J., Marietta, E. V., Miller, M. E. B., Murray, J. A., White, B. A., and Taneja, V., 2012. Loss of Sex and Age Driven Differences in the Gut Microbiome Characterize Arthritis-Susceptible *0401 Mice but Not Arthritis-Resistant *0402 Mice. *PLOS ONE*, 7 (4), e36095.

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., and Ley, R. E., 2016. Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host & Microbe*, 19 (5), 731–743.

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. T., Spector, T. D., Clark, A. G., and Ley, R. E., 2014. Human Genetics Shape the Gut Microbiome. *Cell*, 159 (4), 789–799.

Gorvitovskaia, A., Holmes, S. P., and Huse, S. M., 2016. Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome*, 4 (1), 15.

Goto, Y., 2019. Epithelial Cells as a Transmitter of Signals From Commensal Bacteria and Host Immune Cells. *Frontiers in Immunology* [online], 10. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02057/full [Accessed 22 May 2020].

Govind, N., Choudhury, A., Hodkinson, B., Ickinger, C., Frost, J., Lee, A., Gregersen, P. K., Reynolds, R. J., Bridges, S. L., Hazelhurst, S., Ramsay, M., and Tikly, M., 2014. Immunochip identifies novel, and replicates known, genetic risk loci for rheumatoid arthritis in black South Africans. *Molecular Medicine (Cambridge, Mass.)*, 20, 341–349.

Gregersen, P. K., Amos, C. I., Lee, A. T., Lu, Y., Remmers, E. F., Kastner, D. L., Seldin, M. F., Criswell, L. A., Plenge, R. M., Holers, V. M., Mikuls, T. R., Sokka, T., Moreland, L. W., Bridges, S. L., Xie, G., Begovich, A. B., and Siminovitch, K. A., 2009. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nature Genetics*, 41 (7), 820–823.

Guo, Q., Wang, Y., Xu, D., Nossent, J., Pavlos, N. J., and Xu, J., 2018. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. *Bone Research* [online], 6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5920070/ [Accessed 18 Dec 2019].

Gutzeit, C., Magri, G., and Cerutti, A., 2014. Intestinal IgA production and its role in host-microbe interaction. *Immunological reviews*, 260 (1), 76–85.

Hall, A. B., Tolonen, A. C., and Xavier, R. J., 2017. Human genetic variation and the gut microbiome in disease. *Nature Reviews Genetics* [online], advance online publication. Available from:

http://www.nature.com/nrg/journal/vaop/ncurrent/full/nrg.2017.63.html?WT.mc_id=FBK_NatureReviews [Accessed 21 Aug 2017].

He, J., Li, Y., Cao, Y., Xue, J., and Zhou, X., 2015. The oral microbiome diversity and its relation to human diseases. *Folia Microbiologica*, 60 (1), 69–80.

Hezel, M. P. and Weitzberg, E., 2015. The oral microbiome and nitric oxide homoeostasis. *Oral Diseases*, 21 (1), 7–16.

Hieken, T. J., Chen, J., Hoskin, T. L., Walther-Antonio, M., Johnson, S., Ramaker, S., Xiao, J., Radisky, D. C., Knutson, K. L., Kalari, K. R., Yao, J. Z., Baddour, L. M., Chia, N., and Degnim, A. C., 2016. The Microbiome of Aseptically Collected Human Breast Tissue in Benign and Malignant Disease. *Scientific Reports*, 6, 30751.

Hofer, U., 2014. Microbiome: Pro-inflammatory Prevotella? *Nature Reviews Microbiology*, 12 (1), 5–5.

Hohensinner, P. J., Goronzy, J. J., and Weyand, C. M., 2011. Telomere dysfunction, autoimmunity and aging. *Aging and Disease*, 2 (6), 524–537.

Holers, V. M., Demoruelle, M. K., Kuhn, K. A., Buckner, J. H., Robinson, W. H., Okamoto, Y., Norris, J. M., and Deane, K. D., 2018. Rheumatoid Arthritis and the Mucosal Origins Hypothesis: External Protection Becomes Internal Destruction. *Nature reviews. Rheumatology*, 14 (9), 542–557.

Honne, K., Hallgrímsdóttir, I., Wu, C., Sebro, R., Jewell, N. P., Sakurai, T., Iwamoto, M., Minota, S., and Jawaheer, D., 2016. A longitudinal genome-wide association study of anti-tumor necrosis factor response among Japanese patients with rheumatoid arthritis. *Arthritis Research & Therapy*, 18, 12.

Horta-Baas, G., Romero-Figueroa, M. D. S., Montiel-Jarquín, A. J., Pizano-Zárate, M. L., García-Mena, J., and Ramírez-Durán, N., 2017. Intestinal Dysbiosis and Rheumatoid Arthritis: A Link between Gut Microbiota and the Pathogenesis of Rheumatoid Arthritis. *Journal of Immunology Research*, 2017, 4835189.

Hu, H. J., Jin, E. H., Yim, S. H., Yang, S. Y., Jung, S. H., Shin, S. H., Kim, W. U., Shim, S. C., Kim, T. G., and Chung, Y. J., 2011. Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis. *Experimental & Molecular Medicine*, 43 (11), 613–621.

Hug, H., Mohajeri, M. H., and La Fata, G., 2018. Toll-Like Receptors: Regulators of the Immune Response in the Human Gut. *Nutrients*, 10 (2).

Hugerth, L. W., Wefer, H. A., Lundin, S., Jakobsson, H. E., Lindberg, M., Rodin, S., Engstrand, L., and Andersson, A. F., 2014. DegePrime, a Program for Degenerate Primer Design for Broad-Taxonomic-Range PCR in Microbial Ecology Studies. *Applied and Environmental Microbiology*, 80 (16), 5116–5123.

Hussein, Y. M., Mohamed, R. H., Pasha, H. F., El-Shahawy, E. E., and Alzahrani, S. S., 2011. Association of tumor necrosis factor alpha and its receptor polymorphisms with rheumatoid arthritis in female patients. *Cellular Immunology*, 271 (1), 192–196.

Iljazovic, A., Roy, U., Gálvez, E. J. C., Lesker, T. R., Zhao, B., Gronow, A., Amend, L., Will, S. E., Hofmann, J. D., Pils, M. C., Schmidt-Hohagen, K., Neumann-Schaal, M., and Strowig, T., 2020. Perturbation of the gut microbiome by Prevotella spp . enhances host susceptibility to mucosal inflammation. *Mucosal Immunology*, 1–12.

Jackson, M. A., Bell, J. T., Spector, T. D., and Steves, C. J., 2016. A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units. *PeerJ*, 4, e2341.

Jackson, M. A., Verdi, S., Maxan, M.-E., Shin, C. M., Zierer, J., Bowyer, R. C. E., Martin, T., Williams, F. M. K., Menni, C., Bell, J. T., Spector, T. D., and Steves, C. J., 2018a. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, 9 (1), 2655.

Jackson, M. A., Verdi, S., Maxan, M.-E., Shin, C. M., Zierer, J., Bowyer, R. C. E., Martin, T., Williams, F. M. K., Menni, C., Bell, J. T., Spector, T. D., and Steves, C. J., 2018b. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, 9 (1), 2655.

Jiang, L., Yin, J., Ye, L., Yang, J., Hemani, G., Liu, A.-J., Zou, H., He, D., Sun, L., Zeng, X., Li, Z., Zheng, Y., Lin, Y., Liu, Y., Fang, Y., Xu, J., Li, Y., Dai, S., Guan, J., Jiang, L., Wei, Q., Wang, Y., Li, Y., Huang, C., Zuo, X., Liu, Y., Wu, X., Zhang, L., Zhou, L., Zhang, Q., Li, T., Chen, L., Xu, Z., Yang, X., Qian, F., Xie, W., Liu, W., Guo, Q., Huang, S., Zhao, J., Li, M., Jin, Y., Gao, J., Lv, Y., Wang, Y., Lin, L., Guo, A., Danoy, P., Willner, D., Cremin, C., Hadler, J., Zhang, F., Zhao, Y., Li, M., Yue, T., Fan, X., Guo, J., Mu, R., Li, J., Wu, C., Zeng, M., Wang, J., Li, S., Jin, L., Wang, B., Wang, J., Ma, X., Sun, L., Zhang, X., Brown, M. A., Visscher, P. M., Su, D.-F., and Xu, H., 2014. Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans. *Arthritis & Rheumatology (Hoboken, N.J.)*, 66 (5), 1121–1132.

Jones, E. J., Booth, C., Fonseca, S., Parker, A., Cross, K., Miquel-Clopés, A., Hautefort, I., Mayer, U., Wileman, T., Stentz, R., and Carding, S. R., 2020. The Uptake, Trafficking, and Biodistribution of Bacteroides thetaiotaomicron Generated Outer Membrane Vesicles. *Frontiers in Microbiology* [online], 11. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00057/full [Accessed 24 Feb 2020].

Joo, Y. B., Kim, Y., Park, Y., Kim, K., Ryu, J. A., Lee, S., Bang, S.-Y., Lee, H.-S., Yi, G.-S., and Bae, S.-C., 2017. Biological function integrated prediction of severe radiographic progression in rheumatoid arthritis: a nested case control study. *Arthritis Research & Therapy*, 19 (1), 244.

Joo, Y. B., Park, Y., Kim, K., Bang, S.-Y., Bae, S.-C., and Lee, H.-S., 2018. Association of CD8+ T-cells with bone erosion in patients with rheumatoid arthritis. *International Journal of Rheumatic Diseases*, 21 (2), 440–446.

Julià, A., Ballina, J., Cañete, J. D., Balsa, A., Tornero-Molina, J., Naranjo, A., Alperi-López, M., Erra, A., Pascual-Salcedo, D., Barceló, P., Camps, J., and Marsal, S., 2008. Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. *Arthritis and Rheumatism*, 58 (8), 2275–2286.

Julià, A., Blanco, F., Fernández-Gutierrez, B., González, A., Cañete, J. D., Maymó, J., Alperi-López, M., Olivè, A., Corominas, H., Martínez-Taboada, V., González-Álvaro, I., Fernandez-Nebro, A., Erra, A., Sánchez-Fernández, S., Alonso, A., López-Lasanta, M., Tortosa, R., Codó, L., Lluis Gelpi, J., García-Montero, A. C., Bertranpetit, J., Absher, D., Myers, R. M., Tornero, J., and Marsal, S., 2016. Identification of IRX1 as a Risk

Locus for Rheumatoid Factor Positivity in Rheumatoid Arthritis in a Genome-Wide Association Study. *Arthritis & Rheumatology (Hoboken, N.J.)*, 68 (6), 1384–1391.

Jung, H., Jung, S. M., Rim, Y. A., Park, N., Nam, Y., Lee, J., Park, S.-H., and Ju, J. H., 2017. Arthritic role of Porphyromonas gingivalis in collagen-induced arthritis mice. *PLOS ONE*, 12 (11), e0188698.

Kabat, A. M., Srinivasan, N., and Maloy, K. J., 2014. Modulation of immune development and function by intestinal microbiota. *Trends in Immunology*, 35 (11), 507–517.

Kharlamova, N., Jiang, X., Sherina, N., Potempa, B., Israelsson, L., Quirke, A.-M., Eriksson, K., Yucel-Lindberg, T., Venables, P. J., Potempa, J., Alfredsson, L., and Lundberg, K., 2016. Antibodies to Porphyromonas gingivalis Indicate Interaction Between Oral Infection, Smoking, and Risk Genes in Rheumatoid Arthritis Etiology. *Arthritis & Rheumatology*, 68 (3), 604–613.

Kim, D. and Kim, W., 2016. Editorial: Can Prevotella copri Be a Causative Pathobiont in Rheumatoid Arthritis? *Arthritis & Rheumatology (Hoboken, N.j.)*, 68 (11), 2565–2567.

Kim, K., Bang, S.-Y., Lee, H.-S., and Bae, S.-C., 2017. Update on the genetic architecture of rheumatoid arthritis. *Nature Reviews Rheumatology*, 13 (1), 13–24.

Kim, K., Bang, S.-Y., Lee, H.-S., Cho, S.-K., Choi, C.-B., Sung, Y.-K., Kim, T.-H., Jun, J.-B., Yoo, D. H., Kang, Y. M., Kim, S.-K., Suh, C.-H., Shim, S.-C., Lee, S.-S., Lee, J., Chung, W. T., Choe, J.-Y., Shin, H. D., Lee, J.-Y., Han, B.-G., Nath, S. K., Eyre, S., Bowes, J., Pappas, D. A., Kremer, J. M., Gonzalez-Gay, M. A., Rodriguez-Rodriguez, L., Ärlestig, L., Okada, Y., Diogo, D., Liao, K. P., Karlson, E. W., Raychaudhuri, S., Rantapää-Dahlqvist, S., Martin, J., Klareskog, L., Padyukov, L., Gregersen, P. K., Worthington, J., Greenberg, J. D., Plenge, R. M., and Bae, S.-C., 2015. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Annals of the Rheumatic Diseases*, 74 (3), e13.

Kinane, D. F., Stathopoulou, P. G., and Papapanou, P. N., 2017. Periodontal diseases. *Nature Reviews Disease Primers*, 3 (1), 1–14.

Kiryluk, K., Li, Y., Scolari, F., Sanna-Cherchi, S., Choi, M., Verbitsky, M., Fasel, D., Lata, S., Prakash, S., Shapiro, S., Fischman, C., Snyder, H. J., Appel, G., Izzi, C., Viola, B. F., Dallera, N., Del Vecchio, L., Barlassina, C., Salvi, E., Bertinetto, F. E., Amoroso, A., Savoldi, S., Rocchietti, M., Amore, A., Peruzzi, L., Coppo, R., Salvadori, M., Ravani, P., Magistroni, R., Ghiggeri, G. M., Caridi, G., Bodria, M., Lugani, F., Allegri, L., Delsante, M., Maiorana, M., Magnano, A., Frasca, G., Boer, E., Boscutti, G., Ponticelli, C., Mignani, R., Marcantoni, C., Di Landro, D., Santoro, D., Pani, A., Polci, R., Feriozzi, S., Chicca, S., Galliani, M., Gigante, M., Gesualdo, L., Zamboli, P., Battaglia, G. G., Garozzo, M., Maixnerová, D., Tesar, V., Eitner, F., Rauen, T., Floege, J., Kovacs, T., Nagy, J., Mucha, K., Pączek, L., Zaniew, M., Mizerska-Wasiak, M., Roszkowska-Blaim, M., Pawlaczyk, K., Gale, D., Barratt, J., Thibaudin, L., Berthoux, F., Canaud, G., Boland, A., Metzger, M., Panzer, U., Suzuki, H., Goto, S., Narita, I., Caliskan, Y., Xie, J., Hou, P., Chen, N., Zhang, H., Wyatt, R. J., Novak, J., Julian, B. A., Feehally, J., Stengel, B., Cusi, D., Lifton, R. P., and Gharavi, A. G., 2014. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nature Genetics*, 46 (11), 1187–1196.

Kochi, Y., Okada, Y., Suzuki, A., Ikari, K., Terao, C., Takahashi, A., Yamazaki, K., Hosono, N., Myouzen, K., Tsunoda, T., Kamatani, N., Furuichi, T., Ikegawa, S., Ohmura, K., Mimori, T., Matsuda, F., Iwamoto, T., Momohara, S., Yamanaka, H., Yamada, R., Kubo, M., Nakamura, Y., and Yamamoto, K., 2010. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nature Genetics*, 42 (6), 515–519.

König, J., Siebenhaar, A., Högenauer, C., Arkkila, P., Nieuwdorp, M., Norén, T., Ponsioen, C. Y., Rosien, U., Rossen, N. G., Satokari, R., Stallmach, A., Vos, W. de, Keller, J., and Brummer, R. J., 2017. Consensus report: faecal microbiota transfer – clinical applications and procedures. *Alimentary Pharmacology & Therapeutics*, 45 (2), 222–239.

Konig, M. F., Abusleme, L., Reinholdt, J., Palmer, R. J., Teles, R. P., Sampson, K., Rosen, A., Nigrovic, P. A., Sokolove, J., Giles, J. T., Moutsopoulos, N. M., and Andrade, F., 2016. Aggregatibacter actinomycetemcomitans-induced hypercitrullination links periodontal infection to autoimmunity in rheumatoid arthritis. *Science Translational Medicine*, 8 (369), 369ra176.

Konig, M. F., Paracha, A. S., Moni, M., Bingham, C. O., and Andrade, F., 2015. Defining the role of Porphyromonas gingivalis peptidylarginine deiminase (PPAD) in rheumatoid arthritis through the study of PPAD biology. *Annals of the Rheumatic Diseases*, 74 (11), 2054–2061.

Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R., 2017. Uncultivated microbes in need of their own taxonomy. *The ISME Journal*, 11 (11), 2399–2406.

van der Kraan, P. M., 2017. The changing role of TGFβ in healthy, ageing and osteoarthritic joints. *Nature Reviews Rheumatology*, 13 (3), 155–163.

Krintel, S. B., Palermo, G., Johansen, J. S., Germer, S., Essioux, L., Benayed, R., Badi, L., Ostergaard, M., and Hetland, M. L., 2012. Investigation of single nucleotide polymorphisms and biological pathways associated with response to TNFα inhibitors in patients with rheumatoid arthritis. *Pharmacogenetics and Genomics*, 22 (8), 577–589.

Kulkarni, H. M. and Jagannadham, M. V., 2014. Biogenesis and multifaceted roles of outer membrane vesicles from Gram-negative bacteria. *Microbiology (Reading, England)*, 160 (Pt 10), 2109–2121.

Kurowska, W., Kuca-Warnawin, E. H., Radzikowska, A., and Maśliński, W., 2017. The role of anti-citrullinated protein antibodies (ACPA) in the pathogenesis of rheumatoid arthritis. *Central-European Journal of Immunology*, 42 (4), 390–398.

Lamas, B., Richard, M. L., Leducq, V., Pham, H.-P., Michel, M.-L., Da Costa, G., Bridonneau, C., Jegou, S., Hoffmann, T. W., Natividad, J. M., Brot, L., Taleb, S., Couturier-Maillard, A., Nion-Larmurier, I., Merabtene, F., Seksik, P., Bourrier, A., Cosnes, J., Ryffel, B., Beaugerie, L., Launay, J.-M., Langella, P., Xavier, R. J., and Sokol, H., 2016. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nature Medicine*, 22 (6), 598–605.

Larsen, J. M., 2017. The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology*, 151 (4), 363–374.

Lazar, V., Ditu, L.-M., Pircalabioru, G. G., Gheorghe, I., Curutiu, C., Holban, A. M., Picu, A., Petcu, L., and Chifiriuc, M. C., 2018. Aspects of Gut Microbiota and Immune

System Interactions in Infectious Diseases, Immunopathology, and Cancer. *Frontiers in Immunology* [online], 9. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2018.01830/full [Accessed 18 Jun 2020].

Li, J., Li, Y., Zhou, Y., Wang, C., Wu, B., and Wan, J., 2018. Actinomyces and Alimentary Tract Diseases: A Review of Its Biological Functions and Pathology. *BioMed Research International* [online], 2018. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6129341/ [Accessed 22 Nov 2019].

Li, S., Yu, Y., Yue, Y., Zhang, Z., and Su, K., 2013. Microbial Infection and Rheumatoid Arthritis. *Journal of clinical & cellular immunology* [online], 4 (6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4131749/.

Li, Y., Hingamp, P., Watai, H., Endo, H., Yoshida, T., and Ogata, H., 2018. Degenerate PCR Primers to Reveal the Diversity of Giant Viruses in Coastal Waters. *Viruses* [online], 10 (9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6163766/ [Accessed 18 Aug 2020].

Lie, E., Uhlig, T., van der Heijde, D., Rødevand, E., Kalstad, S., Kaufmann, C., Mikkelsen, K., and Kvien, T. K., 2012. Effectiveness of sulfasalazine and methotrexate in 1102 DMARD-naïve patients with early RA. *Rheumatology*, 51 (4), 670–678.

Lin, L. and Zhang, J., 2017. Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC immunology*, 18 (1), 2.

Lin, R., Zhou, L., Zhang, J., and Wang, B., 2015. Abnormal intestinal permeability and microbiota in patients with autoimmune hepatitis. *International Journal of Clinical and Experimental Pathology*, 8 (5), 5153–5160.

Liu, C., Batliwalla, F., Li, W., Lee, A., Roubenoff, R., Beckman, E., Khalili, H., Damle, A., Kern, M., Furie, R., Dupuis, J., Plenge, R. M., Coenen, M. J. H., Behrens, T. W., Carulli, J. P., and Gregersen, P. K., 2008. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Molecular Medicine (Cambridge, Mass.)*, 14 (9–10), 575–581.

Liu, X., Zou, Q., Zeng, B., Fang, Y., and Wei, H., 2013. Analysis of fecal Lactobacillus community structure in patients with early rheumatoid arthritis. *Current Microbiology*, 67 (2), 170–176.

Lo Presti, A., Zorzi, F., Del Chierico, F., Altomare, A., Cocca, S., Avola, A., De Biasio, F., Russo, A., Cella, E., Reddel, S., Calabrese, E., Biancone, L., Monteleone, G., Cicala, M., Angeletti, S., Ciccozzi, M., Putignani, L., and Guarino, M. P. L., 2019. Fecal and Mucosal Microbiota Profiling in Irritable Bowel Syndrome and Inflammatory Bowel Disease. *Frontiers in Microbiology* [online], 10. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01655/full [Accessed 25 Jul 2019].

Long, D. R., Hsu, J. E., Salipante, S. J., and Bumgarner, R. E., 2019. Letter to the Editor regarding Qui et al: "Cutibacterium acnes and the shoulder microbiome". *Journal of shoulder and elbow surgery*, 28 (8), e275–e276.

Love, M. I., Huber, W., and Anders, S., 2014a. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15 (12), 550.

Love, M. I., Huber, W., and Anders, S., 2014b. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* [online], 15 (12). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302049/ [Accessed 3 Jul 2019].

Lozupone, C. and Knight, R., 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, 71 (12), 8228–8235.

Lundberg, K., Kinloch, A., Fisher, B. A., Wegner, N., Wait, R., Charles, P., Mikuls, T. R., and Venables, P. J., 2008. Antibodies to citrullinated alpha-enolase peptide 1 are specific for rheumatoid arthritis and cross-react with bacterial enolase. *Arthritis and Rheumatism*, 58 (10), 3009–3019.

Lundmark, A., Hu, Y. O. O., Huss, M., Johannsen, G., Andersson, A. F., and Yucel-Lindberg, T., 2019. Identification of Salivary Microbiota and Its Association With Host Inflammatory Mediators in Periodontitis. *Frontiers in Cellular and Infection Microbiology* [online], 9. Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2019.00216/full#h3 [Accessed 12 Jul 2020].

Lustgarten, M. S., 2016. Classifying Aging As a Disease: The Role of Microbes. *Frontiers in Genetics*, 7, 212.

Ma, N. and Ma, X., 2019. Dietary Amino Acids and the Gut-Microbiome-Immune Axis: Physiological Metabolism and Therapeutic Prospects. *Comprehensive Reviews in Food Science and Food Safety*, 18 (1), 221–242.

Macfarlane, S., Bahrami, B., and Macfarlane, G. T., 2011. Mucosal biofilm communities in the human intestinal tract. *Advances in Applied Microbiology*, 75, 111–143.

Maeda, Y., Kurakawa, T., Umemoto, E., Motooka, D., Ito, Y., Gotoh, K., Hirota, K., Matsushita, M., Furuta, Y., Narazaki, M., Sakaguchi, N., Kayama, H., Nakamura, S., Iida, T., Saeki, Y., Kumanogoh, A., Sakaguchi, S., and Takeda, K., 2016. Dysbiosis Contributes to Arthritis Development via Activation of Autoreactive T Cells in the Intestine. *Arthritis & Rheumatology*, 68 (11), 2646–2661.

Maekawa, T., Krauss, J. L., Abe, T., Jotwani, R., Triantafilou, M., Triantafilou, K., Hashim, A., Hoch, S., Curtis, M. A., Nussbaum, G., Lambris, J. D., and Hajishengallis, G., 2014. Porphyromonas gingivalis manipulates complement and TLR signaling to uncouple bacterial clearance from inflammation and promote dysbiosis. *Cell host & microbe*, 15 (6), 768–778.

Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V., and Thiele, I., 2015. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in Genetics* [online], 6. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2015.00148/full [Accessed 8 Apr 2020].

Mandal, R. K., Jiang, T., Al-Rubaye, A. A., Rhoads, D. D., Wideman, R. F., Zhao, J., Pevzner, I., and Kwon, Y. M., 2016. An investigation into blood microbiota and its potential association with Bacterial Chondronecrosis with Osteomyelitis (BCO) in Broilers. *Scientific Reports* [online], 6 (1). Available from: http://www.nature.com/articles/srep25882 [Accessed 9 Mar 2017].

Mankia, K., Cheng, Z., Do, T., Hunt, L., Meade, J., Kang, J., Clerehugh, V., Speirs, A., Tugnait, A., Hensor, E. M. A., Nam, J. L., Devine, D. A., and Emery, P., 2019. Prevalence of Periodontal Disease and Periodontopathic Bacteria in Anti-Cyclic Citrullinated Protein Antibody-Positive At-Risk Adults Without Arthritis. *JAMA network open*, 2 (6), e195394.

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., Quraishi, M. N., Kinross, J., Smidt, H., Tuohy, K. M., Thomas, L. V., Zoetendal, E. G., and Hart, A., 2016. The gut microbiota and host health: a new clinical frontier. *Gut*, 65 (2), 330–339.

Marotte, H., Farge, P., Gaudin, P., Alexandre, C., Mougin, B., and Miossec, P., 2006. The association between periodontal disease and joint destruction in rheumatoid arthritis extends the link between the HLA-DR shared epitope and severity of bone destruction. *Annals of the Rheumatic Diseases*, 65 (7), 905–909.

Márquez, A., Vidal-Bralo, L., Rodríguez-Rodríguez, L., González-Gay, M. A., Balsa, A., González-Álvaro, I., Carreira, P., Ortego-Centeno, N., Ayala-Gutiérrez, M. M., García-Hernández, F. J., González-Escribano, M. F., Sabio, J. M., Tolosa, C., Suárez, A., González, A., Padyukov, L., Worthington, J., Vyse, T., Alarcón-Riquelme, M. E., and Martín, J., 2017. A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, 76 (1), 286–294.

McInnes, I. B. and Schett, G., 2011. The Pathogenesis of Rheumatoid Arthritis. *New England Journal of Medicine*, 365 (23), 2205–2219.

McInnes, I. B. and Schett, G., 2017. Pathogenetic insights from the treatment of rheumatoid arthritis. *The Lancet*, 389 (10086), 2328–2337.

McMurdie, P. J. and Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8 (4), e61217.

McMurdie, P. J. and Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, 10 (4), e1003531.

Menni, C., Zierer, J., Valdes, A. M., and Spector, T. D., 2017. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. *Nature Reviews Rheumatology*, 13 (3), 174–181.

Möller, B., Kollert, F., Sculean, A., and Villiger, P. M., 2020. Infectious Triggers in Periodontitis and the Gut in Rheumatoid Arthritis (RA): A Complex Story About Association and Causality. *Frontiers in Immunology* [online], 11. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2020.01108/full [Accessed 12 Jun 2020].

Mora Ortiz, M., 2019. *(PDF) Exploring host-human microbial interactions in the colon using twins* [online]. ResearchGate. Available from: https://www.researchgate.net/publication/336253445_Exploring_host-human_microbial_interactions_in_the_colon_using_twins?channel=doi&linkId=5d97179f299bf1c363f64066&showFulltext=true [Accessed 21 Aug 2020].

Morton, J. T., Sanders, J., Quinn, R. A., McDonald, D., Gonzalez, A., Vázquez-Baeza, Y., Navas-Molina, J. A., Song, S. J., Metcalf, J. L., Hyde, E. R., Lladser, M., Dorrestein, P.

C., and Knight, R., 2017. Balance Trees Reveal Microbial Niche Differentiation. *mSystems*, 2 (1), e00162-16.

Moss, E. L., Maghini, D. G., and Bhatt, A. S., 2020. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nature Biotechnology*, 38 (6), 701–707.

Mueller, R. B., Kaegi, T., Finckh, A., Haile, S. R., Schulze-Koops, H., and von Kempis, J., 2014. Is radiographic progression of late-onset rheumatoid arthritis different from young-onset rheumatoid arthritis? Results from the Swiss prospective observational cohort. *Rheumatology*, 53 (4), 671–677.

Murugesan, S., Al Ahmad, S. F., Singh, P., Saadaoui, M., Kumar, M., and Al Khodor, S., 2020. Profiling the Salivary microbiome of the Qatari population. *Journal of Translational Medicine*, 18 (1), 127.

Myouzen, K., Kochi, Y., Okada, Y., Terao, C., Suzuki, A., Ikari, K., Tsunoda, T., Takahashi, A., Kubo, M., Taniguchi, A., Matsuda, F., Ohmura, K., Momohara, S., Mimori, T., Yamanaka, H., Kamatani, N., Yamada, R., Nakamura, Y., and Yamamoto, K., 2012. Functional variants in NFKBIE and RTKN2 involved in activation of the NF-κB pathway are associated with rheumatoid arthritis in Japanese. *PLoS genetics*, 8 (9), e1002949.

Negi, S., Juyal, G., Senapati, S., Prasad, P., Gupta, A., Singh, S., Kashyap, S., Kumar, A., Kumar, U., Gupta, R., Kaur, S., Agrawal, S., Aggarwal, A., Ott, J., Jain, S., Juyal, R. C., and Thelma, B. K., 2013. A genome-wide association study reveals ARL15, a novel non-HLA susceptibility gene for rheumatoid arthritis in North Indians. *Arthritis and Rheumatism*, 65 (12), 3026–3035.

Németh, T., Futosi, K., Sitaru, C., Ruland, J., and Mócsai, A., 2016a. Neutrophil-specific deletion of the CARD9 gene expression regulator suppresses autoantibody-induced inflammation *in vivo*. *Nature Communications*, 7, 11004.

Németh, T., Futosi, K., Sitaru, C., Ruland, J., and Mócsai, A., 2016b. Neutrophil-specific deletion of the CARD9 gene expression regulator suppresses autoantibody-induced inflammation in vivo. *Nature Communications*, 7, 11004.

Nishida, A., Inoue, R., Inatomi, O., Bamba, S., Naito, Y., and Andoh, A., 2018. Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clinical Journal of Gastroenterology*, 11 (1), 1–10.

Ogawa, T., Hirose, Y., Honda-Ogawa, M., Sugimoto, M., Sasaki, S., Kibi, M., Kawabata, S., Ikebe, K., and Maeda, Y., 2018. Composition of salivary microbiota in elderly subjects. *Scientific Reports*, 8 (1), 414.

Okada, Y., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Kawaguchi, T., Stahl, E. A., Kurreeman, F. A. S., Nishida, N., Ohmiya, H., Myouzen, K., Takahashi, M., Sawada, T., Nishioka, Y., Yukioka, M., Matsubara, T., Wakitani, S., Teshima, R., Tohma, S., Takasugi, K., Shimada, K., Murasawa, A., Honjo, S., Matsuo, K., Tanaka, H., Tajima, K., Suzuki, T., Iwamoto, T., Kawamura, Y., Tanii, H., Okazaki, Y., Sasaki, T., Gregersen, P. K., Padyukov, L., Worthington, J., Siminovitch, K. A., Lathrop, M., Taniguchi, A., Takahashi, A., Tokunaga, K., Kubo, M., Nakamura, Y., Kamatani, N., Mimori, T., Plenge, R. M., Yamanaka, H., Momohara, S., Yamada, R., Matsuda, F., and Yamamoto, K., 2012. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nature Genetics*, 44 (5), 511–516.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R. R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J. C., Carroll, R. J., Eyler, A. E., Greenberg, J. D., Kremer, J. M., Pappas, D. A., Jiang, L., Yin, J., Ye, L., Su, D.-F., Yang, J., Xie, G., Keystone, E., Westra, H.-J., Esko, T., Metspalu, A., Zhou, X., Gupta, N., Mirel, D., Stahl, E. A., Diogo, D., Cui, J., Liao, K., Guo, M. H., Myouzen, K., Kawaguchi, T., Coenen, M. J. H., van Riel, P. L. C. M., van de Laar, M. A. F. J., Guchelaar, H.-J., Huizinga, T. W. J., Dieudé, P., Mariette, X., Louis Bridges Jr, S., Zhernakova, A., Toes, R. E. M., Tak, P. P., Miceli-Richard, C., Bang, S.-Y., Lee, H.-S., Martin, J., Gonzalez-Gay, M. A., Rodriguez-Rodriguez, L., Rantapää-Dahlqvist, S., Ärlestig, L., Choi, H. K., Kamatani, Y., Galan, P., Lathrop, M., the RACI Consortium, the GARNET Consortium, Eyre, S., Bowes, J., Barton, A., de Vries, N., Moreland, L. W., Criswell, L. A., Karlson, E. W., Taniguchi, A., Yamada, R., Kubo, M., Liu, J. S., Bae, S.-C., Worthington, J., Padyukov, L., Klareskog, L., Gregersen, P. K., Raychaudhuri, S., Stranger, B. E., De Jager, P. L., Franke, L., Visscher, P. M., Brown, M. A., Yamanaka, H., Mimori, T., Takahashi, A., Xu, H., Behrens, T. W., Siminovitch, K. A., Momohara, S., Matsuda, F., Yamamoto, K., and Plenge, R. M., 2014. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, 506 (7488), 376–381.

Oksanen, J., Blanchett, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Scoecs, E., and Wagner, H., 2019. *Package 'Vegan'. Community Ecology Package* [online]. Available from: https://cran.ism.ac.jp/web/packages/vegan/vegan.pdf.

Orchard, T. R., 2012. Management of Arthritis in Patients with Inflammatory Bowel Disease. *Gastroenterology & Hepatology*, 8 (5), 327–329.

Orozco, G., Viatte, S., Bowes, J., Martin, P., Wilson, A. G., Morgan, A. W., Steer, S., Wordsworth, P., Hocking, L. J., UK Rheumatoid Arthritis Genetics Consortium, Wellcome Trust Case Control Consortium, Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate Consortium, Barton, A., Worthington, J., and Eyre, S., 2014. Novel rheumatoid arthritis susceptibility locus at 22q12 identified in an extended UK genome-wide association study. *Arthritis & Rheumatology (Hoboken, N.J.)*, 66 (1), 24–30.

Ortiz, P., Bissada, N. F., Palomo, L., Han, Y. W., Al-Zahrani, M. S., Panneerselvam, A., and Askari, A., 2009. Periodontal therapy reduces the severity of active rheumatoid arthritis in patients treated with or without tumor necrosis factor inhibitors. *Journal of Periodontology*, 80 (4), 535–540.

Padyukov, L., Seielstad, M., Ong, R. T. H., Ding, B., Rönnelid, J., Seddighzadeh, M., Alfredsson, L., Klareskog, L., and Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study group, 2011. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 70 (2), 259–265.

Pahari, S., Chatterjee, D., Negi, S., Kaur, J., Singh, B., and Agrewala, J. N., 2017. Morbid Sequences Suggest Molecular Mimicry between Microbial Peptides and Self-Antigens: A Possibility of Inciting Autoimmunity. *Frontiers in Microbiology*, 8, 1938.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P., 2018. A proposal for a standardized bacterial taxonomy based on genome phylogeny. *bioRxiv*, 256800.

Paul, B., Barnes, S., Demark-Wahnefried, W., Morrow, C., Salvador, C., Skibola, C., and Tollefsbol, T. O., 2015. Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases. *Clinical Epigenetics* [online], 7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4609101/ [Accessed 16 Jan 2017].

Paun, A. and Danska, J. S., 2015. Immuno-ecology: how the microbiome regulates tolerance and autoimmunity. *Current Opinion in Immunology*, 37, 34–39.

Pedersen, H. K., Gudmundsdottir, V., Nielsen, H. B., Hyotylainen, T., Nielsen, T., Jensen, B. A. H., Forslund, K., Hildebrand, F., Prifti, E., Falony, G., Le Chatelier, E., Levenez, F., Doré, J., Mattila, I., Plichta, D. R., Pöhö, P., Hellgren, L. I., Arumugam, M., Sunagawa, S., Vieira-Silva, S., Jørgensen, T., Holm, J. B., Trošt, K., Consortium, M., Kristiansen, K., Brix, S., Raes, J., Wang, J., Hansen, T., Bork, P., Brunak, S., Oresic, M., Ehrlich, S. D., and Pedersen, O., 2016. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*, 535 (7612), 376–381.

Pera, A., Campos, C., López, N., Hassouneh, F., Alonso, C., Tarazona, R., and Solana, R., 2015. Immunosenescence: Implications for response to infection and vaccination in older people. *Maturitas*, 82 (1), 50–55.

Pianta, A., Arvikar, S., Strle, K., Drouin, E. E., Wang, Q., Costello, C. E., and Steere, A. C., 2016. Evidence for Immune Relevance of Prevotella copri, a Gut Microbe, in Patients with Rheumatoid Arthritis. *Arthritis & Rheumatology (Hoboken, N.J.)*.

Picone, P., Nuzzo, D., Caruana, L., Scafidi, V., and Di Carlo, M., 2014. *Mitochondrial Dysfunction: Different Routes to Alzheimer's Disease Therapy* [online]. Oxidative Medicine and Cellular Longevity. Available from: https://www.hindawi.com/journals/omcl/2014/780179/ [Accessed 19 Jun 2020].

Pillai, S., Gopalan, V., and Lam, A. K.-Y., 2017. Review of sequencing platforms and their applications in phaeochromocytoma and paragangliomas. *Critical Reviews in Oncology/Hematology*, 116, 58–67.

Pischon, N., Pischon, T., Kröger, J., Gülmez, E., Kleber, B.-M., Bernimoulin, J.-P., Landau, H., Brinkmann, P.-G., Schlattmann, P., Zernicke, J., Buttgereit, F., and Detert, J., 2008. Association Among Rheumatoid Arthritis, Oral Hygiene, and Periodontitis. *Journal of Periodontology*, 79 (6), 979–986.

Plenge, R. M., Cotsapas, C., Davies, L., Price, A. L., de Bakker, P. I. W., Maller, J., Pe'er, I., Burtt, N. P., Blumenstiel, B., DeFelice, M., Parkin, M., Barry, R., Winslow, W., Healy, C., Graham, R. R., Neale, B. M., Izmailova, E., Roubenoff, R., Parker, A. N., Glass, R., Karlson, E. W., Maher, N., Hafler, D. A., Lee, D. M., Seldin, M. F., Remmers, E. F., Lee, A. T., Padyukov, L., Alfredsson, L., Coblyn, J., Weinblatt, M. E., Gabriel, S. B., Purcell, S., Klareskog, L., Gregersen, P. K., Shadick, N. A., Daly, M. J., and Altshuler, D., 2007. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nature Genetics*, 39 (12), 1477–1482.

Plenge, R. M., Seielstad, M., Padyukov, L., Lee, A. T., Remmers, E. F., Ding, B., Liew, A., Khalili, H., Chandrasekaran, A., Davies, L. R. L., Li, W., Tan, A. K. S., Bonnard, C., Ong, R. T.

H., Thalamuthu, A., Pettersson, S., Liu, C., Tian, C., Chen, W. V., Carulli, J. P., Beckman, E. M., Altshuler, D., Alfredsson, L., Criswell, L. A., Amos, C. I., Seldin, M. F., Kastner, D. L., Klareskog, L., and Gregersen, P. K., 2007. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *The New England Journal of Medicine*, 357 (12), 1199–1209.

Pointon, J. J., Harvey, D., Karaderi, T., Appleton, L. H., Farrar, C., Stone, M. A., Sturrock, R. D., Brown, M. A., and Wordsworth, B. P., 2010. Elucidating the chromosome 9 association with AS; CARD9 is a candidate gene. *Genes and Immunity*, 11 (6), 490–496.

Pritchard, A. B., Crean, S., Olsen, I., and Singhrao, S. K., 2017. Periodontitis, Microbiomes and their Role in Alzheimer's Disease. *Frontiers in Aging Neuroscience* [online], 9. Available from: https://www.frontiersin.org/articles/10.3389/fnagi.2017.00336/full [Accessed 11 Mar 2020].

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L., 2016. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications*, 469 (4), 967–977.

Raychaudhuri, S., Remmers, E. F., Lee, A. T., Hackett, R., Guiducci, C., Burtt, N. P., Gianniny, L., Korman, B. D., Padyukov, L., Kurreeman, F. A. S., Chang, M., Catanese, J. J., Ding, B., Wong, S., van der Helm-van Mil, A. H. M., Neale, B. M., Coblyn, J., Cui, J., Tak, P. P., Wolbink, G. J., Crusius, J. B. A., van der Horst-Bruinsma, I. E., Criswell, L. A., Amos, C. I., Seldin, M. F., Kastner, D. L., Ardlie, K. G., Alfredsson, L., Costenbader, K. H., Altshuler, D., Huizinga, T. W. J., Shadick, N. A., Weinblatt, M. E., de Vries, N., Worthington, J., Seielstad, M., Toes, R. E. M., Karlson, E. W., Begovich, A. B., Klareskog, L., Gregersen, P. K., Daly, M. J., and Plenge, R. M., 2008. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nature Genetics*, 40 (10), 1216–1223.

Reese, A. T. and Dunn, R. R., 2018. Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance. *mBio* [online], 9 (4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6069118/ [Accessed 23 Aug 2020].

Remmers, E. F., Plenge, R. M., Lee, A. T., Graham, R. R., Hom, G., Behrens, T. W., de Bakker, P. I. W., Le, J. M., Lee, H.-S., Batliwalla, F., Li, W., Masters, S. L., Booty, M. G., Carulli, J. P., Padyukov, L., Alfredsson, L., Klareskog, L., Chen, W. V., Amos, C. I., Criswell, L. A., Seldin, M. F., Kastner, D. L., and Gregersen, P. K., 2007. STAT4 and the Risk of Rheumatoid Arthritis and Systemic Lupus Erythematosus. *The New England journal of medicine*, 357 (10), 977–986.

Ricotta, C. and Podani, J., 2017. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity*, 31, 201–205.

Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J. M., Kuruvilla, F., Lagacé, C., Neale, B., Lo, K. S., Schumm, P., Törkvist, L., National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC), United Kingdom Inflammatory Bowel Disease Genetics Consortium, International Inflammatory Bowel Disease Genetics Consortium, Dubinsky, M. C.,

Brant, S. R., Silverberg, M. S., Duerr, R. H., Altshuler, D., Gabriel, S., Lettre, G., Franke, A., D'Amato, M., McGovern, D. P. B., Cho, J. H., Rioux, J. D., Xavier, R. J., and Daly, M. J., 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics*, 43 (11), 1066–1073.

Roager, H. M., Licht, T. R., Poulsen, S. K., Larsen, T. M., and Bahl, M. I., 2014. Microbial Enterotypes, Inferred by the Prevotella-to-Bacteroides Ratio, Remained Stable during a 6-Month Randomized Controlled Diet Intervention with the New Nordic Diet. *Applied and Environmental Microbiology*, 80 (3), 1142–1149.

Rodriguez-Castaño, G. P., Caro-Quintero, A., Reyes, A., and Lizcano, F., 2017. Advances in Gut Microbiome Research, Opening New Strategies to Cope with a Western Lifestyle. *Frontiers in Genetics* [online], 7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5222858/ [Accessed 20 Jun 2020].

de Rooy, D. P. C., Tsonaka, R., Andersson, M. L. E., Forslind, K., Zhernakova, A., Frank-Bertoncelj, M., de Kovel, C. G. F., Koeleman, B. P. C., van der Heijde, D. M. F. M., Huizinga, T. W. J., Toes, R. E. M., Houwing-Duistermaat, J. J., Ospelt, C., Svensson, B., and van der Helm-van Mil, A. H. M., 2015. Genetic Factors for the Severity of ACPA-negative Rheumatoid Arthritis in 2 Cohorts of Early Disease: A Genome-wide Study. *The Journal of Rheumatology*, 42 (8), 1383–1391.

Rosenbaum, J. T. and Asquith, M. J., 2016. The Microbiome: a Revolution in Treatment for Rheumatic Diseases? *Current Rheumatology Reports*, 18 (10), 62.

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., and Tuohy, K., 2018. Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition*, 57 (1), 1–24.

Sabharwal, A., Ganley, K., Miecznikowski, J. C., Haase, E. M., Barnes, V., and Scannapieco, F. A., 2019. The salivary microbiome of diabetic and non-diabetic adults with periodontal disease. *Journal of Periodontology*, 90 (1), 26–34.

Sadanand, S., 2019. Beyond bacteria: studies of other host-associated microorganisms. *Nature Research* [online]. Available from: https://www.nature.com/articles/d42859-019-00013-3 [Accessed 6 Mar 2020].

Sampaio-Maia, B., Caldas, I. M., Pereira, M. L., Pérez-Mongiovi, D., and Araujo, R., 2016. Chapter Four - The Oral Microbiome in Health and Its Implication in Oral and Systemic Diseases. *In*: Sariaslani, S. and Michael Gadd, G., eds. *Advances in Applied Microbiology* [online]. Academic Press, 171–210. Available from: http://www.sciencedirect.com/science/article/pii/S0065216416301095 [Accessed 6 Mar 2020].

Sato, K., Takahashi, N., Kato, T., Matsuda, Y., Yokoji, M., Yamada, M., Nakajima, T., Kondo, N., Endo, N., Yamamoto, R., Noiri, Y., Ohno, H., and Yamazaki, K., 2017. Aggravation of collagen-induced arthritis by orally administered Porphyromonas gingivalis through modulation of the gut microbiota and gut immune system. *Scientific Reports*, 7 (1), 6955.

Saxena, R., Plenge, R. M., Bjonnes, A. C., Dashti, H. S., Okada, Y., Gad El Haq, W., Hammoudeh, M., Al Emadi, S., Masri, B. K., Halabi, H., Badsha, H., Uthman, I. W., Margolin, L., Gupta, N., Mahfoud, Z. R., Kapiri, M., Dargham, S. R., Aranki, G., Kazkaz,

L. A., and Arayssi, T., 2017. A Multinational Arab Genome-Wide Association Study Identifies New Genetic Associations for Rheumatoid Arthritis. *Arthritis & Rheumatology (Hoboken, N.J.)*, 69 (5), 976–985.

Sayers, E., MacGregor, A., and Carding, S. R., 2018. Drug-microbiota interactions and treatment response: Relevance to rheumatoid arthritis. *AIMS Microbiology*, 4 (4), 642–654.

Scher, J. U., Joshua, V., Artacho, A., Abdollahi-Roodsaz, S., Öckinger, J., Kullberg, S., Sköld, M., Eklund, A., Grunewald, J., Clemente, J. C., Ubeda, C., Segal, L. N., and Catrina, A. I., 2016. The lung microbiota in early rheumatoid arthritis and autoimmunity. *Microbiome*, 4 (1), 60.

Scher, J. U., Sczesnak, A., Longman, R. S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo, V., Pamer, E. G., Abramson, S. B., Huttenhower, C., and Littman, D. R., 2013. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*, 2, e01202.

Scher, J. U., Ubeda, C., Equinda, M., Khanin, R., Buischi, Y., Viale, A., Lipuma, L., Attur, M., Pillinger, M. H., Weissmann, G., Littman, D. R., Pamer, E. G., Bretz, W. A., and Abramson, S. B., 2012. Periodontal disease and the oral microbiota in new-onset rheumatoid arthritis. *Arthritis and Rheumatism*, 64 (10), 3083–3094.

Schliep, K. P., 2011. phangorn: phylogenetic analysis in R. *Bioinformatics (Oxford, England)*, 27 (4), 592–593.

Schonland, S. O., Lopez, C., Widmann, T., Zimmer, J., Bryl, E., Goronzy, J. J., and Weyand, C. M., 2003. Premature telomeric loss in rheumatoid arthritis is genetically determined and involves both myeloid and lymphoid cell lineages. *Proceedings of the National Academy of Sciences*, 100 (23), 13471–13476.

Schwenzer, A., Quirke, A.-M., Marzeda, A. M., Wong, A., Montgomery, A. B., Sayles, H. R., Eick, S., Gawron, K., Chomyszyn-Gajewska, M., Łazarz-Bartyzel, K., Davis, S., Potempa, J., Kessler, B. M., Fischer, R., Venables, P. J., Payne, J. B., Mikuls, T. R., and Midwood, K. S., 2017. Association of Distinct Fine Specificities of Anti-Citrullinated Peptide Antibodies With Elevated Immune Responses to Prevotella intermedia in a Subgroup of Patients With Rheumatoid Arthritis and Periodontitis. *Arthritis & Rheumatology (Hoboken, N.J.)*, 69 (12), 2303–2313.

Searle, S. D., Mitnitski, A., Gahbauer, E. A., Gill, T. M., and Rockwood, K., 2008. A standard procedure for creating a frailty index. *BMC Geriatrics*, 8 (1), 24.

Senapati, S., Singh, S., Das, M., Kumar, A., Gupta, R., Kumar, U., Jain, S., Juyal, R. C., and Thelma, B. K., 2014. Genome-wide analysis of methotrexate pharmacogenomics in rheumatoid arthritis shows multiple novel risk variants and leads for TYMS regulation. *Pharmacogenetics and Genomics*, 24 (4), 211–219.

Shah, R. M., McKenzie, E. J., Rosin, M. T., Jadhav, S. R., Gondalia, S. V., Rosendale, D., and Beale, D. J., 2020. An Integrated Multi-Disciplinary Perspectivefor Addressing Challenges of the Human Gut Microbiome. *Metabolites*, 10 (3), 94.

Shan, M., Gentile, M., Yeiser, J. R., Walland, A. C., Bornstein, V. U., Chen, K., He, B., Cassis, L., Bigas, A., Cols, M., Comerma, L., Huang, B., Blander, J. M., Xiong, H., Mayer, L., Berin, C., Augenlicht, L. H., Velcich, A., and Cerutti, A., 2013. Mucus enhances gut

homeostasis and oral tolerance by delivering immunoregulatory signals. *Science (New York, N.Y.)*, 342 (6157), 447–453.

Shannon, C. and Weaver, W., 1949. The Mathematical Theory of Communication, 131.

Sharon, G., Garg, N., Debelius, J., Knight, R., Dorrestein, P. C., and Mazmanian, S. K., 2014. Specialized Metabolites from the Microbiome in Health and Disease. *Cell Metabolism*, 20 (5), 719–730.

Sherina, N., Sippl, N., Israelsson, L., Maitre, E. L., Kharlamova, N., Hansson, M., Eriksson, K., Yucel-Lindberg, T., Malmström, V., Amara, K., and Lundberg, K., 2019. P042/O05 Molecular mimicry and autoimmunity: anti-P.gingivalis antibody response in ACPA-positive rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 78 (Suppl 1), A16–A17.

Shinebaum, R., Neumann, V. C., Cooke, E. M., and Wright, V., 1987. Comparison of Faecal Florae in Patients with Rheumatoid Arthritis and Controls. *British Journal of Rheumatology*, 26, 329–333.

Simpson, E. H., 1949. Measurement of Diversity. *Nature*, 163 (4148), 688–688.

Singh, R. K., Chang, H.-W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., Abrouk, M., Farahnik, B., Nakamura, M., Zhu, T. H., Bhutani, T., and Liao, W., 2017. Influence of diet on the gut microbiome and implications for human health. *Journal of Translational Medicine* [online], 15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5385025/ [Accessed 6 Mar 2020].

Smolen, J. S., Landewé, R. B. M., Bijlsma, J. W. J., Burmester, G. R., Dougados, M., Kerschbaumer, A., McInnes, I. B., Sepriano, A., Vollenhoven, R. F. van, Wit, M. de, Aletaha, D., Aringer, M., Askling, J., Balsa, A., Boers, M., Broeder, A. A. den, Buch, M. H., Buttgereit, F., Caporali, R., Cardiel, M. H., Cock, D. D., Codreanu, C., Cutolo, M., Edwards, C. J., Eijk-Hustings, Y. van, Emery, P., Finckh, A., Gossec, L., Gottenberg, J.-E., Hetland, M. L., Huizinga, T. W. J., Koloumas, M., Li, Z., Mariette, X., Müller-Ladner, U., Mysler, E. F., Silva, J. A. P. da, Poór, G., Pope, J. E., Rubbert-Roth, A., Ruyssen-Witrand, A., Saag, K. G., Strangfeld, A., Takeuchi, T., Voshaar, M., Westhovens, R., and Heijde, D. van der, 2020. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Annals of the Rheumatic Diseases* [online]. Available from: https://ard.bmj.com/content/early/2020/01/22/annrheumdis-2019-216655 [Accessed 17 Apr 2020].

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, R. L., 1998. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*, 25 (2), 134–144.

Staels, B. and Fonseca, V. A., 2009. Bile Acids and Metabolic Regulation. *Diabetes Care*, 32 (Suppl 2), S237–S245.

Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P., Li, Y., Kurreeman, F. A. S., Zhernakova, A., Hinks, A., Guiducci, C., Chen, R., Alfredsson, L., Amos, C. I., Ardlie, K. G., BIRAC Consortium, Barton, A., Bowes, J., Brouwer, E., Burtt, N. P., Catanese, J. J., Coblyn, J., Coenen, M. J. H., Costenbader, K. H., Criswell, L. A., Crusius, J. B. A., Cui, J., de Bakker, P. I. W., De Jager, P. L., Ding, B., Emery, P., Flynn, E., Harrison, P., Hocking, L. J., Huizinga, T. W. J., Kastner, D. L., Ke, X., Lee, A. T., Liu, X., Martin, P., Morgan, A. W., Padyukov, L., Posthumus, M. D., Radstake, T.

R. D. J., Reid, D. M., Seielstad, M., Seldin, M. F., Shadick, N. A., Steer, S., Tak, P. P., Thomson, W., van der Helm-van Mil, A. H. M., van der Horst-Bruinsma, I. E., van der Schoot, C. E., van Riel, P. L. C. M., Weinblatt, M. E., Wilson, A. G., Wolbink, G. J., Wordsworth, B. P., YEAR Consortium, Wijmenga, C., Karlson, E. W., Toes, R. E. M., de Vries, N., Begovich, A. B., Worthington, J., Siminovitch, K. A., Gregersen, P. K., Klareskog, L., and Plenge, R. M., 2010. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genetics*, 42 (6), 508–514.

Steenbergen, H. W. van, Silva, J. A. P. da, Huizinga, T. W. J., and Mil, A. H. M. van der H., 2017. Preventing progression from arthralgia to arthritis: targeting the right patients. *Nature Reviews Rheumatology*, nrrheum.2017.185.

Steer, S. E., Williams, F. M. K., Kato, B., Gardner, J. P., Norman, P. J., Hall, M. A., Kimura, M., Vaughan, R., Aviv, A., and Spector, T. D., 2007. Reduced telomere length in rheumatoid arthritis is independent of disease activity and duration. *Annals of the Rheumatic Diseases*, 66 (4), 476–480.

Stentz, R., Horn, N., Cross, K., Salt, L., Brearley, C., Livermore, D. M., and Carding, S. R., 2015. Cephalosporinases associated with outer membrane vesicles released by Bacteroides spp. protect gut pathogens and commensals against β-lactam antibiotics. *The Journal of Antimicrobial Chemotherapy*, 70 (3), 701–709.

Stentz, R., Osborne, S., Horn, N., Li, A. W. H., Hautefort, I., Bongaerts, R., Rouyer, M., Bailey, P., Shears, S. B., Hemmings, A. M., Brearley, C. A., and Carding, S. R., 2014. A Bacterial Homolog of a Eukaryotic Inositol Phosphate Signaling Enzyme Mediates Cross-kingdom Dialog in the Mammalian Gut. *Cell Reports*, 6 (4), 646–656.

Steves, C. J., Bird, S., Williams, F. M. K., and Spector, T. D., 2016. The Microbiome and Musculoskeletal Conditions of Aging: A Review of Evidence for Impact and Potential Therapeutics. *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research*, 31 (2), 261–269.

Sun, D.-L., Jiang, X., Wu, Q. L., and Zhou, N.-Y., 2013. Intragenomic Heterogeneity of 16S rRNA Genes Causes Overestimation of Prokaryotic Diversity. *Applied and Environmental Microbiology*, 79 (19), 5962–5969.

Svendsen, A. J., Holm, N. V., Kyvik, K., Petersen, P. H., and Junker, P., 2002. Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population. *BMJ (Clinical research ed.)*, 324 (7332), 264–266.

Takeshita, T., Kageyama, S., Furuta, M., Tsuboi, H., Takeuchi, K., Shibata, Y., Shimazaki, Y., Akifusa, S., Ninomiya, T., Kiyohara, Y., and Yamashita, Y., 2016. Bacterial diversity in saliva and oral health-related conditions: the Hisayama Study. *Scientific Reports*, 6 (1), 22164.

Terao, C., Yamada, R., Ohmura, K., Takahashi, M., Kawaguchi, T., Kochi, Y., Human Disease Genomics Working Group, RA Clinical and Genetic Study Consortium, Okada, Y., Nakamura, Y., Yamamoto, K., Melchers, I., Lathrop, M., Mimori, T., and Matsuda, F., 2011. The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Human Molecular Genetics*, 20 (13), 2680–2685.

Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., De Filippis, F., Magnabosco, C., Bonneau, R., Lusingu, J., Amuasi, J., Reinhard, K., Rattei, T., Boulund, F., Engstrand, L., Zink, A., Collado, M. C., Littman, D. R., Eibach, D., Ercolini, D., Rota-Stabelli, O., Huttenhower, C., Maixner, F., and Segata, N., 2019. The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host & Microbe*.

Thaiss, C. A., Zmora, N., Levy, M., and Elinav, E., 2016. The microbiome and innate immunity. *Nature*, 535 (7610), 65–74.

Tian, Y., Zhou, Y., Huang, S., Li, J., Zhao, K., Li, X., Wen, X., and Li, X., 2019. Fecal microbiota transplantation for ulcerative colitis: a prospective clinical study. *BMC Gastroenterology*, 19 (1), 116.

Tsuda, A., Suda, W., Morita, H., Takanashi, K., Takagi, A., Koga, Y., and Hattori, M., 2015. Influence of Proton-Pump Inhibitors on the Luminal Microbiota in the Gastrointestinal Tract. *Clinical and Translational Gastroenterology*, 6 (6), e89.

Ungaro, R., Mehandru, S., Allen, P. B., Peyrin-Biroulet, L., and Colombel, J.-F., 2017. Ulcerative colitis. *The Lancet*, 389 (10080), 1756–1770.

Vaahtovuo, J., Munukka, E., Korkeamäki, M., Luukkainen, R., and Toivanen, P., 2008. Fecal microbiota in early rheumatoid arthritis. *The Journal of Rheumatology*, 35 (8), 1500–1505.

Verdi, S., Abbasian, G., Bowyer, R. C. E., Lachance, G., Yarand, D., Christofidou, P., Mangino, M., Menni, C., Bell, J. T., Falchi, M., Small, K. S., Williams, F. M. K., Hammond, C. J., Hart, D. J., Spector, T. D., and Steves, C. J., 2019. TwinsUK: The UK Adult Twin Registry Update. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 1–7.

Vermilyea, D. M., Ottenberg, G. K., and Davey, M. E., 2019. Citrullination mediated by PPAD constrains biofilm formation in P. gingivalis strain 381. *npj Biofilms and Microbiomes*, 5 (1), 1–11.

Vila, A. V., Collij, V., Sanna, S., Sinha, T., Imhann, F., Bourgonje, A. R., Mujagic, Z., Jonkers, D. M. A. E., Masclee, A. A. M., Fu, J., Kurilshikov, A., Wijmenga, C., Zhernakova, A., and Weersma, R. K., 2020. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nature Communications*, 11 (1), 1–11.

Visconti, A., Roy, C. I. L., Rosa, F., Rossi, N., Martin, T. C., Mohney, R. P., Li, W., Rinaldis, E. de, Bell, J. T., Venter, J. C., Nelson, K. E., Spector, T. D., and Falchi, M., 2019. Interplay between the human gut microbiome and host metabolism. *Nature Communications*, 10 (1), 1–10.

Vos, W. M. de, 2015. Microbial biofilms and the human intestinal microbiome. *npj Biofilms and Microbiomes*, 1 (1), 1–3.

Vuik, F., Dicksved, J., Lam, S., Fuhler, G., van der Laan, L., van de Winkel, A., Konstantinov, S., Spaander, M., Peppelenbosch, M., Engstrand, L., and Kuipers, E., 2019. Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. *United European Gastroenterology Journal*, 7 (7), 897–907.

Wade, W. G. and Prosdocimi, E. M., 2020. Profiling of Oral Bacterial Communities: *Journal of Dental Research* [online]. Available from:

https://journals.sagepub.com/doi/10.1177/0022034520914594 [Accessed 9 Jul 2020].

Wagner, B. D., Grunwald, G. K., Zerbe, G. O., Mikulich-Gilbertson, S. K., Robertson, C. E., Zemanick, E. T., and Harris, J. K., 2018. On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. *Frontiers in Microbiology* [online], 9. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01037/full [Accessed 4 May 2020].

Walsh, A. M., Whitaker, J. W., Huang, C. C., Cherkas, Y., Lamberth, S. L., Brodmerkel, C., Curran, M. E., and Dobrin, R., 2016. Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations. *Genome Biology*, 17, 79.

Wang, J., Bansal, A. T., Martin, M., Germer, S., Benayed, R., Essioux, L., Lee, J. S., Begovich, A., Hemmings, A., Kenwright, A., Taylor, K. E., Upmanyu, R., Cutler, P., Harari, O., Marchini, J., Criswell, L. A., and Platt, A., 2013. Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. *The Pharmacogenomics Journal*, 13 (3), 235–241.

Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., Wei, S., and Zhao, F., 2013. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Scientific Reports*, 3, 1843.

Wasimuddin, Schlaeppi, K., Ronchi, F., Leib, S. L., Erb, M., and Ramette, A., 2019. *Evaluation of primer pairs for microbiome profiling across a food chain from soils to humans within the One Health framework* [online]. Microbiology. preprint. Available from: http://biorxiv.org/lookup/doi/10.1101/843144 [Accessed 6 Jun 2020].

Webster, A. P., Plant, D., Ecker, S., Zufferey, F., Bell, J. T., Feber, A., Paul, D. S., Beck, S., Barton, A., Williams, F. M. K., and Worthington, J., 2018. Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins. *Genome Medicine*, 10 (1), 64.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., and Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5 (1), 27.

Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447 (7145), 661–678.

Weyand, C. M., Yang, Z., and Goronzy, J. J., 2014. T Cell Aging in Rheumatoid Arthritis. *Current opinion in rheumatology*, 26 (1), 93–100.

Wijbrandts, C. A. and Tak, P. P., 2017. Prediction of Response to Targeted Treatment in Rheumatoid Arthritis. *Mayo Clinic Proceedings*, 92 (7), 1129–1143.

Wilsdon, T. D. and Hill, C. L., 2017. Managing the drug treatment of rheumatoid arthritis. *Australian Prescriber*, 40 (2), 51–58.

Wilson, C., Tiwana, H., and Ebringer, A., 2000. Molecular mimicry between HLA-DR alleles associated with rheumatoid arthritis and Proteus mirabilis as the Aetiological basis for autoimmunity. *Microbes and Infection*, 2 (12), 1489–1496.

Wu, H.-J., Ivanov, I. I., Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D. R., Benoist, C., and Mathis, D., 2010. Gut-Residing Segmented Filamentous Bacteria Drive Autoimmune Arthritis via T Helper 17 Cells. *Immunity*, 32 (6), 815–827.

Wu, X., He, B., Liu, J., Feng, H., Ma, Y., Li, D., Guo, B., Liang, C., Dang, L., Wang, L., Tian, J., Zhu, H., Xiao, L., Lu, C., Lu, A., and Zhang, G., 2016. Molecular Insight into Gut Microbiota and Rheumatoid Arthritis. *International Journal of Molecular Sciences* [online], 17 (3). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4813281/ [Accessed 6 Feb 2017].

Wu, Y., Chi, X., Zhang, Q., Chen, F., and Deng, X., 2018. Characterization of the salivary microbiome in people with obesity. *PeerJ* [online], 6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5858547/ [Accessed 15 Jul 2020].

Yang, B., Wang, Y., and Qian, P.-Y., 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, 17, 135.

Yang, X.-K., Liu, J., Liu, J., Liang, Y., Xu, W.-D., Leng, R.-X., Pan, H.-F., and Ye, D.-Q., 2015. Associations Between PADI4 Gene Polymorphisms and Rheumatoid Arthritis: An Updated Meta-analysis. *Archives of Medical Research*, 46 (4), 317–325.

Yu, F. Y., Wang, Q. Q., Li, M., Cheng, Y.-H., Cheng, Y.-S. L., Zhou, Y., Yang, X., Zhang, F., Ge, X., Zhao, B., and Ren, X. Y., 2020. Dysbiosis of saliva microbiome in patients with oral lichen planus. *BMC Microbiology*, 20 (1), 1–12.

Zakharzhevskaya, N. B., Vanyushkina, A. A., Altukhov, I. A., Shavarda, A. L., Butenko, I. O., Rakitina, D. V., Nikitina, A. S., Manolov, A. I., Egorova, A. N., Kulikov, E. E., Vishnyakov, I. E., Fisunov, G. Y., and Govorun, V. M., 2017a. Outer membrane vesicles secreted by pathogenic and nonpathogenic Bacteroides fragilis represent different metabolic activities. *Scientific Reports*, 7 (1), 5008.

Zakharzhevskaya, N. B., Vanyushkina, A. A., Altukhov, I. A., Shavarda, A. L., Butenko, I. O., Rakitina, D. V., Nikitina, A. S., Manolov, A. I., Egorova, A. N., Kulikov, E. E., Vishnyakov, I. E., Fisunov, G. Y., and Govorun, V. M., 2017b. Outer membrane vesicles secreted by pathogenic and nonpathogenic Bacteroides fragilis represent different metabolic activities. *Scientific Reports*, 7 (1), 5008.

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W., Tang, S., Xu, X., Wang, X., Cai, X., Liu, S., Xia, Y., Li, J., Qiao, X., Al-Aama, J. Y., Chen, H., Wang, L., Wu, Q., Zhang, F., Zheng, W., Li, Y., Zhang, M., Luo, G., Xue, W., Xiao, L., Li, J., Chen, W., Xu, X., Yin, Y., Yang, H., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Y., and Wang, J., 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Medicine*, 21 (8), 895–905.

Zhou, Y. and Zhi, F., 2016. *Lower Level of Bacteroides in the Gut Microbiota Is Associated with Inflammatory Bowel Disease: A Meta-Analysis* [online]. BioMed Research International. Available from: https://www.hindawi.com/journals/bmri/2016/5828959/ [Accessed 5 Jul 2020].

Zhu, W., Winter, M. G., Byndloss, M. X., Spiga, L., Duerkop, B. A., Hughes, E. R., Büttner, L., Romão, E. de L., Behrendt, C. L., Lopez, C. A., Sifuentes-Dominguez, L., Huff-Hardy, K., Wilson, R. P., Gillis, C. C., Tükel, Ç., Koh, A. Y., Burstein, E., Hooper, L. V.,

Bäumler, A. J., and Winter, S. E., 2018. Precision editing of the gut microbiota ameliorates colitis. *Nature*, 553 (7687), 208–211.

Appendices

Appendix A: Chapter 3

Script for generation of ASVs in TwinsUK

```
source("http://bioconductor.org/biocLite.R")
biocLite(suppressUpdates = FALSE)
biocLite("ShortRead", suppressUpdates = FALSE)
biocLite("devtools")
library("devtools")
biocLite("dada2")
remove.packages("ade4")
install.packages("ade4")
rm(list=ls())
remove.packages("ape")
install.packages("ape")
update.packages("ape")
library(dada2)
```

```
#Demultiplex in Qiime specifying RC barcodes
split_libraries_fastq.py -i Undetermined_S0_L001_R1_001.fastq.gz -o High2R1Dmplxed/ -b
Undetermined_S0_L001_I1_001.fastq.gz -m High2Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
```

```
split_libraries_fastq.py -i Undetermined_S0_L001_R2_001.fastq.gz -o High2R2Dmplxed/ -b
Undetermined_S0_L001_I1_001.fastq.gz -m High2Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o High1R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o High1R2split/ --file_type fastq
#AllHigh1
split_libraries_fastq.py -i Undetermined_S0_L001_R1_001.fastq.gz -o AllHigh1R1Dmplxed/ -b
Undetermined_S0_L001_I1_001.fastq.gz -m AllHigh1Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsHigh1R1.fastq -o AllHigh1R1split/ --file_type fastq
 source /macqiime/configs/bash_profile.txt
#Dada2
path <- file.path("~/Documents/ASVs2/TwinsUKHigh1/R1R2")
list.files(path)
fnFs <- sort(list.files(path, pattern="R1", full.names = TRUE))
fnRs <- sort(list.files(path, pattern="R2", full.names = TRUE))
plotQualityProfile(fnFs)
plotQualityProfile(fnRs)
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)ge

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
```

```
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)

#High1

setwd("~/Documents/ASVs2/High1/AllHigh1")

#Dada2
path <- file.path("AllHigh1R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("AllHigh1R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/High1/AllHigh1")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
```

```
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.High1.rds")

#High2
#Qiime
Documents/ASVs2/TwinsUKHigh2
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o High2R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o High2R2split/ --file_type fastq
#Desktop
path <- file.path("High2R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("High2R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/TwinsUKHigh2/High2Filt")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
```

```
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.High2.rds")

#Low1
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow1/Undetermined_
S0_L001_R1_001.fastq.gz -o Low1R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow1/Undetermined_
S0_L001_I1_001.fastq.gz -m Low1Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r
999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow1/Undetermined_
S0_L001_R2_001.fastq.gz -o Low1R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow1/Undetermined_
S0_L001_I1_001.fastq.gz -m Low1Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r
999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o Low1R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o Low1R2split/ --file_type fastq
setwd("~/Documents/ASVs2/Low1")

#Desktop
path <- file.path("Low1R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Low1R2split")
list.files(path)
```

```
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Low1")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.Low1.rds")

#Low2
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow2/Undetermined_
S0_L001_R1_001.fastq.gz -o Low2R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow2/Undetermined_
```

```
S0_L001_I1_001.fastq.gz -m Low2Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r
999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow2/Undetermined_
S0_L001_R2_001.fastq.gz -o Low2R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKLow2/Undetermined_
S0_L001_I1_001.fastq.gz -m Low2Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r
999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o Low2R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o Low2R2split/ --file_type fastq

setwd("~/Documents/ASVs2/Low2")

#Desktop
path <- file.path("Low2R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Low2R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6], title(main = Low2R1))
plotQualityProfile(fnRs[1:6], title(main = Low2R1))
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Low2")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
```

```
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.Low2.rds")



#Plate5
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate5/Undetermined
_S0_L001_R1_001.fastq.gz -o Plate5R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate5/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate5Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate5/Undetermined
_S0_L001_R2_001.fastq.gz -o Plate5R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate5/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate5Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o Plate5R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o Plate5R2split/ --file_type fastq

setwd("~/Documents/ASVs2/Plate5")

#Dada2
path <- file.path("Plate5R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate5R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
```

```
plotQualityProfile(fnFs[1:6], title(main = Low2R1))
plotQualityProfile(fnRs[1:6], title(main = Low2R1))
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate5")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.Plate5.rds")

#Plate 6
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate6/Undetermined
_S0_L001_R1_001.fastq.gz -o Plate6R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate6/Undetermined
```

```
_S0_L001_I1_001.fastq.gz -m Plate6Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate6/Undetermined
_S0_L001_R2_001.fastq.gz -o Plate6R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate6/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate6Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o Plate6R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o Plate6R2split/ --file_type fastq
#Dada2
setwd("~/Documents/ASVs2/Plate6")
path <- file.path("Plate6R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate6R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate6")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(180,190),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)


plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)
```

```
names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.Plate6.rds")


#Plate 7
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate7/Undetermined
_S0_L001_R1_001.fastq.gz -o Plate7R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate7/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate7Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate7/Undetermined
_S0_L001_R2_001.fastq.gz -o Plate7R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate7/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate7Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o Plate7R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o Plate7R2split/ --file_type fastq
#Dada2
setwd("~/Documents/ASVs2/Plate7")
path <- file.path("Plate7R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate7R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
```

```
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate7")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)
saveRDS(seqtab, "seqtab.Plate7.rds")



#Plate 8
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate8/Undetermined
_S0_L001_R1_001.fastq.gz -o Plate8R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate8/Undetermined
```

```
_S0_L001_I1_001.fastq.gz -m Plate8Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate8/Undetermined
_S0_L001_R2_001.fastq.gz -o Plate8R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate8/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate8Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001

#Dada2
setwd("~/Documents/ASVs2/Plate8")
path <- file.path("Plate8R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate8R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate8")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
```

```
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))



saveRDS(seqtab, "seqtab.Plate8.rds")
#Plate 9
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate9/Undetermined
_S0_L001_R1_001.fastq.gz -o Plate9R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate9/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate9Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate9/Undetermined
_S0_L001_R2_001.fastq.gz -o Plate9R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate9/Undetermined
_S0_L001_I1_001.fastq.gz -m Plate9Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq
-r 999 -n 999 -q 0 -p 0.0001
split_sequence_file_on_sample_ids.py -i seqsR1.fastq -o Plate9R1split/ --file_type fastq
split_sequence_file_on_sample_ids.py -i seqsR2.fastq -o Plate9R2split/ --file_type fastq

library("dada2")
setwd("~/Documents/ASVs2/Plate9")
path <- file.path("Plate9R1split")
list.files(path)

fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate9R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
```

```
path <- file.path("Documents/ASVs2/Plate9")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(190,190),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)

errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate9.rds")
########

setwd("~/Documents/ASVs2/Plate9")

#Dada2
path <- file.path("Plate9R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate9R2split")
```

```
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:10])
plotQualityProfile(fnFs[10:20])
plotQualityProfile(fnFs[21:31])
plotQualityProfile(fnFs[32:42])
plotQualityProfile(fnFs[43:53])
plotQualityProfile(fnFs[54:64])
plotQualityProfile(fnFs[65:75])
plotQualityProfile(fnFs[75:79])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate9")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(170,170),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)

saveRDS(seqtab, "seqtab.Plate9.rds")
#### copied again
```

```
setwd("~/Documents/ASVs2/Plate9")
path <- file.path("Plate9R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate9R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:10])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("~/Documents/ASVs2/Plate9")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,200),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
```

```
saveRDS(seqtab, "seqtab.Plate9.rds")

 #Plate9 big data PE workflow method


library(dada2); packageVersion("dada2")
setwd("~/Documents/ASVs2/Plate9")

# File parsing
pathF <- "~/Documents/ASVs2/Plate9/Plate9R1split"
list.files(pathF)


pathR <- "~/Documents/ASVs2/Plate9/Plate9R2split"
list.files(pathR)

filtpathF <- file.path(pathF, "filtered")
filtpathF
filtpathR <- file.path(pathR, "filtered")
filtpathR
fastqFs <- sort(list.files(pathF, pattern="fastq"))

fastqRs <- sort(list.files(pathR, pattern="fastq"))

if(length(fastqFs) != length(fastqRs)) stop("Forward and reverse files do not match.")

# Filtering: THESE PARAMETERS ARENT OPTIMAL FOR ALL DATASETS
out <- filterAndTrim(fwd=file.path(pathF, fastqFs), filt=file.path(filtpathF, fastqFs),
            rev=file.path(pathR, fastqRs), filt.rev=file.path(filtpathR, fastqRs), trimLeft = c(12,12),
            truncLen=c(160,160), maxEE=2, truncQ=11, maxN=0, rm.phix=TRUE,
            compress=TRUE, verbose=TRUE, multithread=TRUE)

# File parsing
filtpathF <- "~/Documents/ASVs2/Plate9/Plate9R1split/filtered"
filtpathR <- "~/Documents/ASVs2/Plate9/Plate9R2split/filtered"
filtFs <- list.files(filtpathF, pattern="fastq", full.names = TRUE)
filtRs <- list.files(filtpathR, pattern="fastq", full.names = TRUE)
plotQualityProfile(filtFs[1:4])
sample.namesR <- sample.names
names(filtFs) <- sample.names
names(filtRs) <- sample.names
```

```
set.seed(100)
# Learn forward error rates
errF <- learnErrors(filtFs, nbases=1e8, multithread=TRUE)
# Learn reverse error rates
errR <- learnErrors(filtRs, nbases=1e8, multithread=TRUE)
# Sample inference and merger of paired-end reads
mergers <- vector("list", length(sample.names))
names(mergers) <- sample.names
for(sam in sample.names) {
  cat("Processing:", sam, "\n")
  derepF <- derepFastq(filtFs[[sam]])
  ddF <- dada(derepF, err=errF, multithread=TRUE)
  derepR <- derepFastq(filtRs[[sam]])
  ddR <- dada(derepR, err=errR, multithread=TRUE)
  merger <- mergePairs(ddF, derepF, ddR, derepR)
  mergers[[sam]] <- merger
}

# Construct sequence table and remove chimeras
seqtab <- makeSequenceTable(mergers)
setwd("~/Documents/ASVs2/Plate9")
saveRDS(seqtab, "Plate9BD.rds")



#Plate 10
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate10/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate10R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate10/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate10Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate10/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate10R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate10/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate10Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/Plate10")
path <- file.path("Plate10R1split")
```

```
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate10R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate10")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate10.rds")

#Plate11
```

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate11/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate11R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate11/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate11Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate11/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate11R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate11/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate11Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/Plate11")
path <- file.path("Plate11R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate11R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate11")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)
```

```
names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate11.rds")

#Plate12
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate12/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate12R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate12/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate12Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate12/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate12R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate12/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate12Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/Plate12")
path <- file.path("Plate12R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate12R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate12")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
```

```
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate12.rds")

#Plate13
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate13/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate13R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate13/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate13Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate13/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate13R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate13/Undetermine
```

```
d_S0_L001_I1_001.fastq.gz -m Plate13Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/Plate13")
path <- file.path("Plate13R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate13R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate13")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
```

```
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate13.rds")

#Plate14
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate14R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate14Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate14R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate14Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

#Corrected Barcodes
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate14R1Dmplxed_CorrectedBarcodes/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate14Map_CorrectedBarcodes.txt --barcode_type 12  --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate14R2Dmplxed_CorrectedBarcodes/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate14/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate14Map_CorrectedBarcodes.txt --barcode_type 12  --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/Plate14")
path <- file.path("Plate14R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate14R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
```

```
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate14")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate14.rds")

#Plate15
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate15/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate15R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate15/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate15Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
```

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate15/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate15R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate15/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate15Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001


setwd("~/Documents/ASVs2/Plate15")
path <- file.path("Plate15R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate15R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate15")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,200),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
```

```
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate15.rds")
```

#Plate16

```
## Extract barcodes from FASTA files

library(readr)
Plate16 <- read_delim("~/Documents/Plate16.csv",
                      " ", escape_double = FALSE, col_names = FALSE,
                      trim_ws = TRUE)

Plate16$X6 <- gsub("bc_diffs", "", Plate16$X6)
Plate16$X6 <- gsub("=", "", Plate16$X6)
Plate16$X5 <- gsub("new_bc=", "", Plate16$X5)
Plate16$X4 <- gsub("orig_bc=", "", Plate16$X5)
plate16copy <- Plate16
Plate16 <- Plate16[c(1,4:6)]


Plate16 <- as.data.frame(Plate16)
names <- c("Sample", "Orig_Barcode", "New_Barcode", "Changed")
colnames(Plate16) <- names


odd_indexes <- seq(1,1048575,2)
head(odd_indexes)
Plate16sub <- Plate16[c(odd_indexes),]
Plate16sub$Sample <- gsub(">", "", Plate16sub$Sample)
x <- sub("\\_[0-9]+", "", Plate16sub$Sample)
head(x)


Plate16sub$Sample <- x
```

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate16R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate16Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate16R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate16Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
#corrected barcodes
```

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_R1_001.fastq.gz -o Plate16R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate16Map_CorrectedBarcodes.txt --barcode_type 12  --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_R2_001.fastq.gz -o Plate16R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_1/illumina_run_data/TwinsUKPlate16/Undetermine
d_S0_L001_I1_001.fastq.gz -m Plate16Map_CorrectedBarcodes.txt --barcode_type 12  --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

library(dada2)
setwd("~/Documents/ASVs2/Plate16")
path <- file.path("Plate16R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate16R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/Plate16")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)

plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
```

```
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate16.rds")

                                ############################## Batch2
#####################################
#Plate17_21
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate17_21/Undeter
mined_S0_L001_R1_001.fastq.gz -o Plate17_21R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate17_21/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate17_21Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate17_21/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate17_21R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate17_21/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate17_21Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/ReadsBatch2/Plate17_21")
path <- file.path("Plate17_21R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate17_21R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
```

```
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/ReadsBatch2/Plate17_21")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate17_21.rds")

#Plate18_31

split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate18_31/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate18_31R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate18_31/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate18_31Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
```

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate18_31/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate18_31R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate18_31/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate18_31Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/ReadsBatch2/Plate18_31")
path <- file.path("Plate18_31R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate18_31R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/ReadsBatch2/Plate18_31")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
```

```
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate18_31.rds")

#Plate19_30_RD2
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate19_30_RD2/Un
determined_S0_L001_R1_001.fastq.gz -o Plate19_30_RD2R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate19_30_RD2/Un
determined_S0_L001_I1_001.fastq.gz -m Plate19_30_RD2Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate19_30_RD2/Un
determined_S0_L001_R2_001.fastq.gz -o Plate19_30_RD2R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate19_30_RD2/Un
determined_S0_L001_I1_001.fastq.gz -m Plate19_30_RD2Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/ReadsBatch2/Plate19_30_RD2")
path <- file.path("Plate19_30RD2R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate19_30_RD2R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/ReadsBatch2/Plate19_30_RD2")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
outdf <- as.data.frame(out)
write.csv(outdf, file = "FilterandTrim.csv")
```

```
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate19_30_RD2.rds")

#Plate20_23
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_R1_001.fastq.gz -o Plate20_23R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate20_23Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate20_23R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate20_23Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/ReadsBatch2/Plate20_23")
path <- file.path("Plate20_23RD2R1split")
```

```
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate20_23RD2R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/ReadsBatch2/Plate20_23")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
outdf <- as.data.frame(out)
write.csv(outdf, file = "FilterandTrim.csv")
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate20_23_RD2.rds")
```

#Plate22

split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_R1_001.fastq.gz -o Plate20_23R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate20_23Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate20_23R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate20_23/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate20_23Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

#Plate24_29

split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate24_29/Undeter
mined_S0_L001_R1_001.fastq.gz -o Plate24_29R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate24_29/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate24_29Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate24_29/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate24_29R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate24_29/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate24_29Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

#Plate25_27

split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate25_27/Undeter
mined_S0_L001_R1_001.fastq.gz -o Plate25_27R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate25_27/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate25_27Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate25_27/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate25_27R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate25_27/Undeter

```
mined_S0_L001_I1_001.fastq.gz -m Plate25_27Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

#Plate26_RD1
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate26_RD1/Undete
rmined_S0_L001_R1_001.fastq.gz -o Plate26_RD1R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate26_RD1/Undete
rmined_S0_L001_I1_001.fastq.gz -m Plate26_RD1Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate26_RD1/Undete
rmined_S0_L001_R2_001.fastq.gz -o Plate26_RD1R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate26_RD1/Undete
rmined_S0_L001_I1_001.fastq.gz -m Plate26_RD1Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

#Plate28_33
#R1 in wrong format:"Non-header line passed as input. Header must start with '@'."
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate28_33/Undeter
mined_S0_L001_R1_001.fastq.gz -o Plate28_33R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate28_33/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate28_33Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate28_33/Undeter
mined_S0_L001_R2_001.fastq.gz -o Plate28_33R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate28_33/Undeter
mined_S0_L001_I1_001.fastq.gz -m Plate28_33Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

setwd("~/Documents/ASVs2/ReadsBatch2/Plate28_33")


path <- file.path("Plate28_33R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))

plotQualityProfile(fnRs[1:9])
plotQualityProfile(fnRs[10:18])
```

```
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/ReadsBatch2/Plate28_83")

filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim( fnRs, filtRs, trimLeft = 12, truncLen= 160,
                      maxN=0, maxEE= 2, truncQ=2, rm.phix=TRUE,
                      compress=TRUE, multithread=TRUE)
outdf <- as.data.frame(out)
write.csv(outdf, file = "FilterandTrim160.csv")

plotQualityProfile(filtRs[1:4])
head(out)

errR <- learnErrors(filtRs, multithread=TRUE)

plotErrors(errR, nominalQ=TRUE)

#Dereplication

derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepRs) <- sample.names
#Denoising

dadaRs <- dada(derepRs, err=errR, multithread=TRUE)

dadaRs[[2]]

seqtab <- makeSequenceTable(dadaRs)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate28_33.rds")

#Plate32_34_RD3 ..
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate32_34_RD3/Un
determined_S0_L001_R1_001.fastq.gz -o Plate32_34_RD3R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate32_34_RD3/Un
determined_S0_L001_I1_001.fastq.gz -m Plate32_34_RD3Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
```

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate32_34_RD3/Un
determined_S0_L001_R2_001.fastq.gz -o Plate32_34_RD3R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_2/illumina_run_data/TwinsUKPlate32_34_RD3/Un
determined_S0_L001_I1_001.fastq.gz -m Plate32_34_RD3Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001

#########################Batch3###############################
#Plate34.2_40
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate34.2_40/Undetermined_S0_L001_
R1_001.fastq.gz -o Plate34.2_40R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate34.2_40/Undetermined_S0_L001_I
1_001.fastq.gz -m Plate34.2_40Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r
999 -n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate34.2_40/Undetermined_S0_L001_
R2_001.fastq.gz -o Plate34.2_40R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate34.2_40/Undetermined_S0_L001_I
1_001.fastq.gz -m Plate34.2_40Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r
999 -n 999 -q 0 -p 0.0001
setwd("~/Documents/ASVs2/ReadsBatch3/Plate34.2_40")
require(dada2)
#####POOLED = TRUE


library(dada2)
path <- file.path("Plate34.2_40R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate34.2_40R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnFs), "_"), `[`, 1)
sample.names <- gsub(".fastq", "", sample.names)
path <- file.path("Documents/ASVs2/ReadsBatch3/Plate34.2_40")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
```

```
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(200,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
outdf <- as.data.frame(out)
write.csv(outdf, file = "FilterandTrimPOOLED.csv")
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, pool = TRUE, err=errF,  multithread=TRUE)
dadaRs <- dada(derepRs, pool = TRUE, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))

saveRDS(seqtab, "seqtab.Plate46_247POOLED.rds")



#Plate35_36
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate35_36/Undetermined_S0_L001_R
1_001.fastq.gz -o Plate35_36R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate35_36/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate35_36Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate35_36/Undetermined_S0_L001_R
```

2_001.fastq.gz -o Plate35_36R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate35_36/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate35_36Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

#Plate37_38
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate37_38/Undetermined_S0_L001_R
1_001.fastq.gz -o Plate37_38R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate37_38/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate37_38Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate37_38/Undetermined_S0_L001_R
2_001.fastq.gz -o Plate37_38R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate37_38/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate37_38Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

#Plate39_42
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate39_42/Undetermined_S0_L001_R
1_001.fastq.gz -o Plate39_42R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate39_42/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate39_42Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate39_42/Undetermined_S0_L001_R
2_001.fastq.gz -o Plate39_42R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate39_42/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate39_42Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

#Plate41_43
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate41_43/Undetermined_S0_L001_R
1_001.fastq.gz -o Plate41_43R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate41_43/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate41_43Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

```
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate41_43/Undetermined_S0_L001_R
2_001.fastq.gz -o Plate41_43R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate41_43/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate41_43Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

#Plate44_45
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate44_45/Undetermined_S0_L001_R
1_001.fastq.gz -o Plate44_45R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate44_45/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate44_45Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate44_45/Undetermined_S0_L001_R
2_001.fastq.gz -o Plate44_45R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate44_45/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate44_45Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

#Plate46_47
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate46_47/Undetermined_S0_L001_R
1_001.fastq.gz -o Plate46_47R1Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate46_47/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate46_47Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001
split_libraries_fastq.py -i
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate46_47/Undetermined_S0_L001_R
2_001.fastq.gz -o Plate46_47R2Dmplxed/ -b
/gpfs/home/DTR/ubiome/Cornell/gut16s/Data/raw_data/batch_3/TwinsUKPlate46_47/Undetermined_S0_L001_I1
_001.fastq.gz -m Plate46_47Map.txt --barcode_type 12 --rev_comp_barcode --store_demultiplexed_fastq -r 999 -
n 999 -q 0 -p 0.0001

split_libraries_fastq.py -i Undetermined_S0_L001_R1_001.fastq.gz -o Plate46_47R1Dmplxed/ -b
Undetermined_S0_L001_I1_001.fastq.gz -m Plate46_47Map.txt --barcode_type 12 --rev_comp_barcode --
store_demultiplexed_fastq -r 999 -n 999 -q 0 -p 0.0001
library(dada2)

setwd("~/Documents/ASVs2/ReadsBatch3/Plate46_47")
```

```
path <- file.path("Plate46_47R1split")
list.files(path)
fnFs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
path <- file.path("Plate46_47R2split")
list.files(path)
fnRs <- sort(list.files(path, pattern="fastq", full.names = TRUE))
plotQualityProfile(fnFs[1:6])
plotQualityProfile(fnRs[1:6])
sample.names <- sapply(strsplit(basename(fnRs), "_"), `[`, 1)
path <- file.path("Documents/ASVs2/ReadsBatch3/Plate46_47")

filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path,"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, trimLeft = c(12,12), truncLen=c(240,160),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
outdf <- as.data.frame(out)
write.csv(outdf, file = "FilterandTrim.csv")
plotQualityProfile(filtFs[1:4])
plotQualityProfile(filtRs[1:4])
head(out)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)

#Dereplication
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)

names(derepFs) <- sample.names
names(derepRs) <- sample.names
#Denoising
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
dadaFs[[1]]
dadaRs[[2]]
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
```

```
saveRDS(seqtab, "seqtab.Plate46_247.rds")


                                    ############ Merge Seqtabs Batch1 ###################
st1 <- readRDS("~/Documents/ASVs2/High1/AllHigh1/seqtab.High1.rds")
st2 <- readRDS("~/Documents/ASVs2/High2/seqtab.High2.rds")
st3 <- readRDS("~/Documents/ASVs2/Low1/seqtab.Low1.rds")
st4 <- readRDS("~/Documents/ASVs2/Low2/seqtab.Low2.rds")
st5 <- readRDS("~/Documents/ASVs2/Plate5/seqtab.Plate5.rds")
st6 <- readRDS("~/Documents/ASVs2/Plate6/seqtab.Plate6.rds")
st7 <- readRDS("~/Documents/ASVs2/Plate7/seqtab.Plate7.rds")
st8 <- readRDS("~/Documents/ASVs2/Plate8/seqtab.Plate8.rds")
st9 <- readRDS("~/Documents/ASVs2/Plate9/Plate9BD.rds")
st10 <- readRDS("~/Documents/ASVs2/Plate10/seqtab.Plate10.rds")
st11 <- readRDS("~/Documents/ASVs2/Plate11/seqtab.Plate11.rds")
st12 <- readRDS("~/Documents/ASVs2/Plate12/seqtab.Plate12.rds")
st13 <- readRDS("~/Documents/ASVs2/Plate13/seqtab.Plate13.rds")
st14 <- readRDS("~/Documents/ASVs2/Plate14/seqtab.Plate14.rds")
st15 <- readRDS("~/Documents/ASVs2/Plate15/seqtab.Plate15.rds")
st16 <- readRDS("~/Documents/ASVs2/Plate16/seqtab.Plate16.rds")


mergeTest <- mergeSequenceTables(st3, st4, st7, st8, st9)
library(dada2)
st.batch1 <- mergeSequenceTables(st1, st2, st3, st4, st5, st6, st7, st8, st9, st10, st11, st12, st13, st14, st15,
st16)


setwd("~/Documents/ASVs2")

saveRDS(st.batch1, file = "st.batch1.rds")

                ############################### Merge Batch 2
########################################

st1 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate17_21/seqtab.Plate17_21.rds")
st2 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate18_31/seqtab.Plate18_31.rds")
st3 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate19_30_RD2/seqtab.Plate19_30_RD2.rds")
st4 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate20_23/seqtab.Plate20_23_RD2.rds")
st5 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate22/seqtab.Plate22.rds")
st6 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate24_29/seqtab.Plate24_29.rds")
```

```
st7 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate25_27/seqtab.Plate25_27.rds")
st8 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate26_RD1/seqtab.Plate26_RD1.rds")
st9 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate28_33/seqtab.Plate28_33.rds")
st10 <- readRDS("~/Documents/ASVs2/ReadsBatch2/Plate32_34/seqtab.Plate32_34.rds")

st.batch2 <- mergeSequenceTables(st1, st2, st3, st4, st5, st6, st7, st8, st9, st10)

setwd("~/Documents/ASVs2")
saveRDS(st.batch2, file = "st.batch2.rds")


                ############################## Merge Batch 3
########################################
st1 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate34.2_40/seqtab.Plate34.2_40.rds")
st2 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate35_36/seqtab.Plate35_36.rds")
st3 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate37_38/seqtab.Plate37_38.rds")
st4 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate39_42/seqtab.Plate39_42.rds")
st5 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate41_43/seqtab.Plate41_43.rds")
st6 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate44_45/seqtab.Plate44_45.rds")
st7 <- readRDS("~/Documents/ASVs2/ReadsBatch3/Plate46_47/seqtab.Plate46_247.rds")

st.batch3 <- mergeSequenceTables(st1, st2, st3, st4, st5, st6, st7)

############################## Merge All ######################################
st.batch1 <- readRDS("~/Documents/ASVs2/st.batch1.rds")

merged12 <- mergeSequenceTables(st.batch1, st.batch2)
merged23 <- mergeSequenceTables(st.batch2, st.batch3)
merged123 <- mergeSequenceTables(merged23, st.batch1)


TwinsUkSeqtab <- merged123

saveRDS(TwinsUkSeqtab, file = "TwinsUKSeqtab.rds")

seqtab.nochim <- removeBimeraDenovo(TwinsUkSeqtab, method="consensus", multithread=TRUE,
verbose=TRUE)
sum(seqtab.nochim)/sum(TwinsUkSeqtab)
setwd("~/Documents/ASVs2")
saveRDS(seqtab.nochim, file = "TwinsUKSeqtab.nochim.rds")
```

```
#Check Merge has worked
seqtab.nochim.df <- as.data.frame(seqtab.nochim)
write.csv(seqtab.nochim.df, file = "TwinsUKSeqTabNoChim.csv")
write.table(seqtab.nochim.df, file = "TwinsUKSeqTabNoChim.txt")
st3nochim <- removeBimeraDenovo(st3)
st3nochim <- as.data.frame(st3nochim)
st3nochim$Total <- NA
st3nochim$Total <- rowSums(st3nochim)
st3nochimtotals <- st3nochim[(4363)]
seqtab.nochim.df$Total <- NA
seqtab.nochim.df$Total <- rowSums(seqtab.nochim.df[c(1:36699)])
seqtab.nochim.totals <- seqtab.nochim.df[(36700)]
seqtab.nochim.500 <- seqtab.nochim.df[which(seqtab.nochim.df$Total > 500),]
st5nochim <- removeBimeraDenovo(st5)

TwinsUKstnochim <- readRDS("~/Documents/ASVs2/TwinsUKSeqtab.nochim.rds")
ASVIDs <- read.csv("~/Documents/ASVs2/ASV_IDs.csv")
TwinsUKstnochimsub <- TwinsUKstnochim[which(rownames(TwinsUKstnochim) %in% ASVIDs$ASV_ID),]
subcopy <- TwinsUKstnochimsub
row.names(TwinsUKstnochimsub) <- ASVIDs$SequencingSpecificName
saveRDS( TwinsUKstnochimsub,file = "TwinsUKASVs_Depth10000.rds")
seqtab.nochim.df <- as.data.frame(TwinsUKstnochimsub)

write.csv(seqtab.nochim.df, file = "TwinsUK_ASVs_10000.csv")
write.table(seqtab.nochim.df, file = "TwinsUK_ASVs_10000.txt", sep = "\t")

library(dada2)
seqs <- readRDS("TwinsUKASVs_Depth10000.rds")
taxa <- assignTaxonomy(seqs, "silva_nr_v132_train_set.fa.gz")
taxa.sp <- addSpecies(taxa, "silva_species_assignment_v132.fa.gz")
saveRDS(taxa.sp, file = "TwinsUK_Taxa.rds", )
write.csv(taxa.sp, file = "TwinsUK_Taxa.csv")
saveRDS(taxa.sp, file = "TwinsUK_Taxa.rds")

ST2 <- readRDS("~/Documents/ASVs2/TwinsUKASVs_Depth10000.rds")



ASVs <-
readRDS("~/Documents/ASVs2/FinalData/Depth10000_DeError_Decontam/TwinsUK_ASVsDepth10000_NoCon
tam.rds")
ASVs <-collapseNoMismatch(ASVs, minOverlap = 20, orderBy = "abundance",
```

```
                identicalOnly = FALSE, vec = TRUE, verbose = FALSE)
ASVs <- as.data.frame(ASVs)

setwd("~/Documents/ASVs2/FinalData/Depth10000_DeError_Decontam/Collapsed")

saveRDS(ASVs, file = "ASVsCollapsed.rds")

#Taxa Assign
taxa <- assignTaxonomy(seqtab.nochim,"silva_nr_v132_train_set.fa.gz", multithread=TRUE)
unname(taxa)
#Species Assign
taxa.sp <- addSpecies(taxa, "silva_species_assignment_v132.fa.gz", verbose=TRUE)
unname(taxa.sp)

Library/Frameworks/R.framework/Versions/3.4/Resources/library
#PHLYLOGENETIC TREE
#Set to run for a week on cluster
source("https://bioconductor.org/biocLite.R")
biocLite("DECIPHER")
require(DECIPHER)
seqs <- getSequences(seqtab.nochim)
names(seqs) <- seqs # This propagates to the tip labels of the tree
alignment <- AlignSeqs(DNAStringSet(seqs), anchor=NA,verbose=FALSE)

require(phangorn)
phangAlign <- phyDat(as(alignment, "matrix"), type="DNA")
dm <- dist.ml(phangAlign)
# Note, tip order != sequence order
treeNJ <- NJ(dm)
fit = pml(treeNJ, data=phangAlign)
fitGTR <- update(fit, k=4, inv=0.2)
fitGTR <- optim.pml(fitGTR, model="GTR", optInv=TRUE, optGamma=TRUE,
                rearrangement = "stochastic", control = pml.control(trace = 0))
detach("package:phangorn", unload=TRUE)
Tree1 <- fitGTR$tree
write.tree(Tree1, file = "ASVsTree.tre")
```

Appendix B Chapter 5: Supplementary material

The following supplementary materials are associated with the paper published in *The Lancet Rheumatology* on 1st July 2020, entitled "Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross sectional study".

Supplementary Methods

	SE Positive (n=74)	SE Negative (n=55)	T Test
Age Median (IQR)	56.5 (12.75)	57 (18.5)	0.69
Female Sex n (%)	62 (84)	47 (85.5)	0.79
BMI Median (IQR)	24.5 (4.8)	24.4 (4.22)	0.47
ACPA positivity n (%)	20 (27)	16 (29)	0.79

Supplementary Table B 5.1. Age, sex, BMI and ACPA positivity in SE positive and SE negative groups in the SCREEN-RA cohort

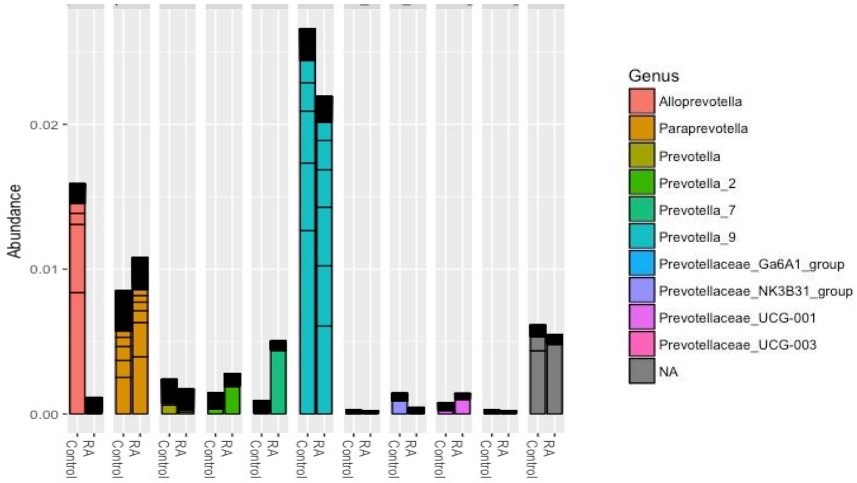
Resilience hypothesis: abundance of *Prevotella spp*. in RA discordant twin pairs in TwinsUK cohort.

To be confident that the *Prevotella spp* findings in non-RA individuals with genetic risk for RA were not reflecting a "resilience-associated" microbiota (i.e. protective factors against developing disease despite having high genetic risk), we calculated the abundance of these species in 19 monozygotic RA discordant female twin pairs from TwinsUK, which had been excluded from the previous analysis because of the presence of disease. If associated with resilience we would expect the *Prevotella spp* to be preferentially expressed in the unaffected twin of each discordant pair. RA diagnosis had been confirmed during a clinical visit to a Rheumatologist or RA-trained research assistant. ASVs were generated as above and relative abundance of *Prevotella spp.* in RA discordant twin pairs calculated using the Phyloseq R package[19] (**Supplementary Figure 1**).

Supplementary Results

Alpha diversity	Estimate	Std. Error	R^2	P
Observed	-0.011	0.021	0.251	0.6
Shannon	-0.007	0.024	0.026	0.76
Simpson	-0.023	0.034	0.006	0.41
Faith's	-0.009	0.021	0.116	0.65

Supplementary Table B 5.2. Association of genetic risk for RA (PRS) and alpha diversity of the gut microbiota in TwinsUK participants.



Supplementary Figure B 5.1. Relative abundance of *Prevotella spp.* in RA-discordant twin pairs from TwinsUK. The RA affected twins were established (treated) RA patients. Prevotella_7 was more abundant in RA affected co-twins, compared to non-RA co-twins.

CHR	SNP	A1	A2	P	OR	Replicated
1	rs624988	T	C	8.00E-10	1.09	No
1	rs840016	C	T	2.00E-06	1.11	No
1	rs923880	C	T	8.00E-06	0.25	No
1	rs72717009	T	C	5.00E-07	1.12	Yes
1	rs2317230	T	G	2.00E-07	1.07	No
1	rs6593803	T	C	2.00E-07	3.13	No
1	rs2228145	A	C	4.00E-09	1.08	Yes
1	rs12565755	C	T	5.00E-08	0.39	No
1	rs2105325	C	A	0	1.12	Yes
1	rs12140275	A	T	4.00E-09	1.11	Yes
1	rs3890745	T	C	1.00E-07	0.89	Yes
1	rs28411352	T	C	0	1.11	Yes
1	rs2301888	G	A	0	1.19	Yes
1	rs12131057	G	A	4.00E-07	1.16	No
1	rs2476601	A	G	2.00E-11	1.72	Yes
1	rs6679677	A	C	0	1.79	Yes
1	rs11121380	C	A	5.00E-08	11.11	No
1	rs998731	C	T	3.00E-09	1.11	Yes
2	rs6732565	A	G	9.00E-09	1.1	Yes
2	rs11676922	T	A	0	1.12	Yes
2	rs10865035	A	G	2.00E-06	1.12	No
2	rs9653442	C	T	0	1.11	Yes
2	rs11900673	T	C	1.00E-08	1.11	No
2	rs13385025	A	G	9.00E-07	1.11	No
2	rs6715284	G	C	3.00E-09	1.15	Yes
2	rs1980422	C	T	0	1.12	Yes
2	rs3087243	G	A	1.00E-08	1.15	Yes
2	rs231735	G	T	6.00E-09	1.17	No
2	rs12617656	T	C	0	1.557	Yes
2	rs41005	T	C	4.00E-06	0.27	No
2	rs1406428	T	C	2.00E-07	2.74	No
2	rs10175798	A	G	1.00E-09	1.08	Yes
2	rs13031237	T	G	8.00E-07	1.13	No
2	rs13017599	A	G	0	1.21	No
2	rs34695944	C	T	0	1.13	Yes
2	rs4305317	T	A	2.00E-06	1.45	No
2	rs13015080	A	G	3.00E-06	0.28	Yes
2	rs934734	G	A	5.00E-10	1.13	Yes
2	rs1858037	T	A	0	1.11	Yes
2	rs7574865	T	G	2.00E-06	1.17	Yes
2	rs11889341	T	C	0	1.13	Yes

CHR	SNP	A1	A2	P	OR	Replicated
3	rs73081554	T	C	5.00E-08	1.18	No
3	rs2062583	G	T	2.00E-06	0.63	No
3	rs9860428	C	A	4.00E-07	1.6	Yes
3	rs3806624	G	A	9.00E-09	1.08	Yes
3	rs9826828	A	G	9.00E-10	1.44	No
3	rs6774280	C	T	9.00E-07	5.5	No
3	rs4452313	T	A	5.00E-11	1.11	Yes
3	rs7624766	G	A	4.00E-07	2.28	No
3	rs13315591	C	T	5.00E-08	1.29	No
3	rs1809529	C	T	3.00E-07	0.28	No
4	rs2867461	A	G	0	1.13	No
4	rs11933540	C	T	0	1.15	No
4	rs13142500	C	T	2.00E-06	1.1	Yes
4	rs13119723	A	G	7.00E-07	1.12	No
4	rs45475795	G	A	4.00E-06	1.14	No
4	rs11937061	T	G	2.00E-07	6.7	No
4	rs3816587	C	T	9.00E-06	1.09	No
4	rs874040	C	G	0	1.19	No
4	rs6448432	A	G	4.00E-07	1.19	No
4	rs13137105	G	A	9.00E-07	6.39	No
4	rs2664035	A	G	3.00E-08	1.08	Yes
5	rs7731626	G	A	0	1.21	No
5	rs255758	C	A	7.00E-06	1.42	No
5	rs26232	C	T	4.00E-08	0.88	No
5	rs2561477	G	A	2.00E-11	1.11	Yes
5	rs1498103	A	T	6.00E-06	0.26	No
5	rs12519788	G	A	5.00E-08	1.82	Yes
5	rs657075	A	G	3.00E-10	1.12	Yes
5	rs6859219	C	A	1.00E-11	1.28	No
5	rs12109285	C	T	1.00E-07	130.2	No
5	rs1502644	C	A	1.00E-08	2.33	No
5	rs4336372	C	T	1.00E-06	0.65	Yes
5	rs4921283	A	G	1.00E-07	1.0752687	No
6	rs6568431	A	C	1.00E-07	1.14	No
6	rs9372120	G	T	8.00E-10	1.11	Yes
6	rs3093024	A	G	0	1.19	No
6	rs3093023	A	G	2.00E-11	1.11	Yes
6	rs1571878	C	T	0	1.17	Yes
6	rs1854853	A	G	0	1.4164	Yes
6	rs12529514	C	T	2.00E-08	1.14	No
6	rs3763309	A	C	0	2.3	No

CHR	SNP	A1	A2	P	OR	Replicated
6	rs6457617	T	C	1.00E-09	2.36	Yes
6	rs9275406	T	G	0	2.1	No
6	rs660895	G	A	0	3.62	No
6	rs7765379	G	T	0	2.51	No
6	rs13192471	G	A	0	1.97	No
6	rs6910071	G	A	0	2.88	No
6	rs9378815	C	G	2.00E-10	1.09	Yes
6	rs284515	G	A	7.00E-07	0.35	No
6	rs284511	C	T	3.00E-08	0.41	No
6	rs2233434	G	A	0	1.19	Yes
6	rs2233424	T	C	0	1.24	Yes
6	rs9373594	T	C	3.00E-09	1.09	Yes
6	rs2451258	T	C	3.00E-11	1.1	Yes
6	rs6920220	A	G	0	1.22	Yes
6	rs5029924	T	C	0	1.74	No
6	rs7752903	G	T	0	1.38	Yes
6	rs10499194	C	T	1.00E-09	1.33	No
6	rs2230926	C	A	2.00E-06	1.31	No
6	rs75908454	C	T	6.00E-07	0.42	No
7	rs42041	G	C	4.00E-06	1.11	No
7	rs4272	G	A	5.00E-09	1.1	Yes
7	rs6956740	T	C	5.00E-07	3.49	No
7	rs11761231	T	C	4.00E-07	1.44	No
7	rs10488631	C	T	4.00E-11	1.16	No
7	rs3807306	T	G	3.00E-07	1.44	No
7	rs67250450	T	C	3.00E-09	1.11	Yes
7	rs12531711	G	A	0	1.45	No
8	rs1600249	T	A	5.00E-06	0.77	No
8	rs2736340	T	C	1.00E-11	1.21	Yes
8	rs2736337	C	T	0	1.11	Yes
8	rs678347	G	A	7.00E-09	1.1	Yes
8	rs16938910	T	C	4.00E-07	28.66	No
8	rs800586	A	G	2.00E-07	9.78	No
8	rs1516971	T	C	3.00E-11	1.16	Yes
9	rs11574914	A	G	0	1.13	Yes
9	rs2812378	G	A	3.00E-08	1.12	No
9	rs951005	A	G	4.00E-10	1.19	No
9	rs12379034	G	A	0	1.489	Yes
9	rs1329568	T	G	8.00E-07	19.64	No
9	rs9299346	A	G	2.00E-06	2.76	No
9	rs2072438	T	C	3.00E-09	1.142	No

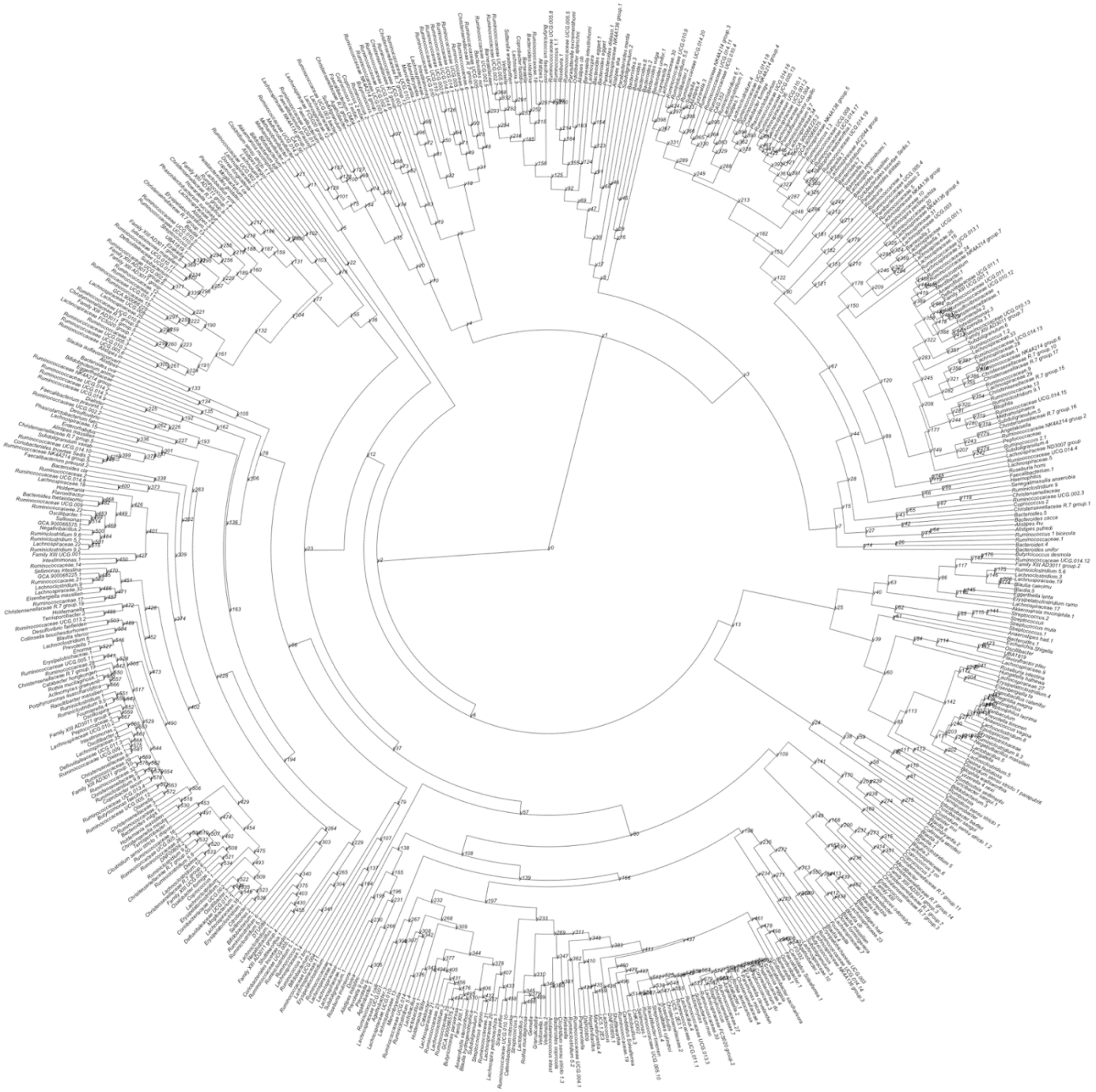
CHR	SNP	A1	A2	P	OR	Replicated
9	rs3761847	G	A	0	1.32	Yes
9	rs10985070	C	A	5.00E-11	1.08	Yes
10	rs10821944	G	T	0	1.16	No
10	rs71508903	T	C	0	1.16	Yes
10	rs3824660	C	T	3.00E-09	1.1	Yes
10	rs1679568	A	G	8.00E-07	0.35	Yes
10	rs706778	T	C	1.00E-11	1.14	Yes
10	rs2104286	C	T	1.00E-06	1.19	No
10	rs12413578	C	T	8.00E-08	1.2	Yes
10	rs16906916	C	A	8.00E-07	14	No
10	rs12570744	T	C	2.00E-07	0.49	No
10	rs4750316	G	C	4.00E-06	0.91	Yes
10	rs947474	A	G	3.00E-10	1.12	Yes
10	rs3125734	T	C	5.00E-09	1.2	No
10	rs6479800	C	G	4.00E-08	1.11	Yes
10	rs726288	T	C	9.00E-09	1.22	No
10	rs2671692	A	G	3.00E-09	1.07	No
10	rs793108	T	C	1.00E-09	1.08	Yes
11	rs508970	A	G	3.00E-06	1.07	No
11	rs4409785	C	T	1.00E-11	1.12	Yes
11	rs10790268	G	A	0	1.17	Yes
11	rs73013527	C	T	1.00E-10	1.09	Yes
11	rs4937362	T	C	8.00E-07	1.09	No
11	rs968567	C	T	2.00E-08	1.12	No
11	rs518167	A	G	2.00E-06	2.24	No
11	rs7940423	A	G	1.00E-07	10.9	No
11	rs72991	C	T	5.00E-07	2.45	No
11	rs1079467	C	G	2.00E-06	0.29	No
11	rs3781913	T	G	6.00E-10	1.12	No
11	rs331463	T	A	1.00E-07	1.12	Yes
12	rs773125	A	G	1.00E-10	1.09	Yes
12	rs1633360	T	C	7.00E-08	1.08	Yes
12	rs11051970	T	G	1.00E-06	2.08	No
12	rs1678542	C	G	1.00E-07	1.2	Yes
12	rs11045392	C	T	2.00E-06	3.3	No
12	rs3184504	T	C	6.00E-06	1.08	No
12	rs10774624	G	A	7.00E-09	1.09	No
12	rs3794271	G	A	4.00E-06	3.2	No
12	rs12831974	C	T	6.00E-06	1.27	No
13	rs9557321	C	T	6.00E-08	1.73	No
13	rs9603612	G	C	0	1.098901	No

CHR	SNP	A1	A2	P	OR	Replicated
13	rs9603616	C	T	0	1.1	Yes
13	rs4942242	T	C	2.00E-07	0.13	No
13	rs9604529	G	A	7.00E-07	0.35	No
13	rs3790022	A	G	1.00E-06	1.4925	No
14	rs2582532	C	T	5.00E-08	1.18	Yes
14	rs7155603	G	A	1.00E-07	1.16	No
14	rs3783637	C	T	2.00E-06	1.1	No
14	rs17118552	G	A	2.00E-07	17.19	No
14	rs1624005	G	A	7.00E-06	2.59	No
14	rs7141276	A	G	2.00E-06	0.37	No
14	rs2841277	T	C	0	1.15	No
14	rs1957895	G	T	4.00E-07	1.09	No
14	rs3783782	A	G	2.00E-09	1.14	Yes
14	rs1950897	T	C	8.00E-11	1.1	Yes
15	rs1914816	A	G	7.00E-07	0.87	No
15	rs1898036	C	T	2.00E-07	7.73	No
15	rs7164176	G	A	5.00E-07	0.28	No
15	rs17374222	A	C	2.00E-06	1.13	No
15	rs8026898	A	G	0	1.15	Yes
15	rs12901682	A	C	4.00E-08	135.02	No
15	rs8032939	C	T	0	1.13	Yes
15	rs6496667	A	C	1.00E-06	1.09	No
16	rs6500395	C	T	6.00E-07	2.2	No
16	rs2280381	T	C	2.00E-06	1.12	No
16	rs13330176	A	T	9.00E-09	1.12	Yes
16	rs16973500	T	NA	9.00E-06	0.24	No
16	rs7404928	T	C	4.00E-06	1.08	No
16	rs4780401	T	G	9.00E-09	1.09	Yes
17	rs72634030	A	C	2.00E-09	1.12	No
17	rs2872507	A	G	9.00E-07	1.1	No
17	rs1877030	C	T	2.00E-08	1.09	No
17	rs12601925	G	A	6.00E-09	2.15	Yes
18	rs2469434	C	T	9.00E-10	1.07	Yes
18	rs16977065	T	C	1.00E-07	33.72	No
18	rs2847297	G	A	2.00E-08	1.1	No
18	rs8083786	G	A	0	1.14	Yes
18	rs6506569	T	C	7.00E-06	1.98	No
18	rs2002842	A	C	6.00E-06	1.15	No
18	rs1943199	T	G	2.00E-08	1.94	No
19	rs7258015	C	T	3.00E-08	1.18	No
19	rs1273516	A	G	9.00E-07	4.26	No

CHR	SNP	A1	A2	P	OR	Replicated
19	rs1543922	T	C	3.00E-07	34.48	No
19	rs34536443	G	C	0	1.46	No
20	rs4810485	G	T	8.00E-09	0.91	No
20	rs4239702	C	T	0	1.13	Yes
20	rs6138892	C	T	3.00E-07	10.93	No
20	rs6026990	A	T	6.00E-07	0.63	No
21	rs2075876	A	G	4.00E-09	1.18	No
21	rs73194058	C	A	3.00E-08	1.13	Yes
21	rs2837960	G	T	2.00E-06	1.05	No
21	rs8133843	A	G	3.00E-09	1.08	Yes
21	rs11203203	A	G	4.00E-06	1.11	No
21	rs1893592	A	C	1.00E-11	1.11	No
22	rs1043099	C	G	7.00E-09	1.19	No
22	rs3218251	A	T	6.00E-06	1.08	No
22	rs743777	G	A	1.00E-06	1.09	No
22	rs909685	A	T	0	1.13	Yes
22	rs2283790	G	A	2.00E-11	1.22	No
22	rs11089637	C	T	2.00E-09	1.08	Yes

Supplementary Table B 5.3. SNPs included in the RA PRS

Appendix C Chapter 6: Dendogram of balances and association with RA



Supplementary Figure C 6.1 Bifurcating tree representing hierarchical clustering of balances generated from core taxa of the gut microbiota within MZ and DZ twins.

Supplementary Table C 6.1 Linear mixed effects model regression of balances against RA affected status

Balance	Group Var	Intercept	P Value
y0	0.999999985	0.852520069	0.804
y1	0.999999985	0.638942962	0.749
y2	0.999999985	0.480894514	0.704
y3	0.999999985	0.355119254	0.903
y4	0.999999985	0.25838904	0.946
y5	0.999999985	0.021725476	0.991
y6	0.999999985	0.818757673	0.897
y7	0.999999985	0.87228188	0.957
y8	0.999999985	0.737304608	0.889
y9	0.999999985	0.491406545	0.892
y10	0.999999985	0.254022443	0.880
y11	0.999999985	0.622370767	0.862
y12	0.999999985	0.219717164	0.852
y13	0.999999985	0.995229955	0.802
y14	0.999999985	0.232248395	0.900
y15	0.999999985	0.70735578	0.876
y16	0.999999985	0.100714002	0.983
y17	0.999999985	0.399402751	0.946
y18	0.999999985	0.198908573	0.979
y19	0.999999985	0.261221776	0.848
y20	0.999999985	0.970211529	0.880
y21	0.999999985	0.234234396	0.743
y22	0.999999985	0.50035056	0.846
y23	0.999999985	0.478952133	0.999
y24	0.999999985	0.588414081	0.937
y25	0.999999985	0.942471228	0.987
y26	0.999999985	0.332971889	0.855
y27	0.999999985	0.651182165	0.706
y28	0.999999985	0.84275485	0.991
y29	0.999999985	0.862377013	0.914
y30	0.999999985	0.090252627	0.857
y31	0.999999985	0.800756544	0.796
y32	0.999999985	0.390069663	0.951
y33	0.999999985	0.602792961	0.818
y34	0.999999985	0.788313513	0.835

y35	0.999999985	0.607343463	0.856
y36	0.999999985	0.557454664	0.958
y37	0.999999985	0.51429241	0.924
y38	0.999999985	0.139660051	0.944
y39	0.999999985	0.511750302	0.986
y40	0.999999985	0.247707705	0.785
y41	0.999999985	0.195378008	0.780
y42	0.999999985	0.987374381	0.854
y43	0.999999985	0.163616928	0.887
y44	0.999999985	0.9914404	0.950
y45	0.999999985	0.357884317	0.912
y46	0.999999985	0.693371548	0.825
y47	0.999999985	0.866993313	0.858
y48	0.999999985	0.874090116	0.929
y49	0.999999985	0.748956894	0.883
y50	0.999999985	0.58697904	0.853
y51	0.999999985	0.122299915	0.970
y52	0.999999985	0.245603286	0.971
y53	0.999999985	0.922976647	0.944
y54	0.999999985	0.459169118	0.791
y55	0.999999985	0.889052026	0.903
y56	0.999999985	0.106543898	0.978
y57	0.999999985	0.924815755	0.860
y58	0.999999985	0.633185972	0.771
y59	0.999999985	0.629318689	0.869
y60	0.999999985	0.022721808	0.991
y61	0.999999985	0.562270869	0.724
y62	0.999999985	0.728952462	0.910
y63	0.999999985	0.747688876	0.859
y64	0.999999985	0.7260767	0.911
y65	0.999999985	0.45016624	0.959
y66	0.999999985	0.135698786	0.969
y67	0.999999985	0.792240093	0.835
y68	0.999999985	0.997371273	0.908
y69	0.999999985	0.800056221	0.970
y70	0.999999985	0.74900135	0.969
y71	0.999999985	0.209472947	0.999
y72	0.999999985	0.84337477	0.994
y73	0.999999985	0.764618972	0.940
y74	0.999999985	0.718795072	0.939
y75	0.999999985	0.975394704	0.938
y76	0.999999985	0.982423273	0.959

y77	0.999999985	0.433677221	0.701
y78	0.999999985	0.987848531	0.897
y79	0.999999985	0.242649409	0.871
y80	0.999999985	0.983147607	0.980
y81	0.999999985	0.711867284	0.926
y82	0.999999985	0.742064028	0.879
y83	0.999999985	0.512758945	0.866
y84	0.999999985	0.794190205	0.887
y85	0.999999985	0.385075209	0.784
y86	0.999999985	0.124345016	0.997
y87	0.999999985	0.222718778	0.904
y88	0.999999985	0.727573043	0.962
y89	0.999999985	0.969545783	0.964
y90	0.999999985	0.849333805	0.968
y91	0.999999985	0.736687581	0.964
y92	0.999999985	0.892025613	0.993
y93	0.999999985	0.935850991	0.826
y94	0.999999985	0.28547213	0.951
y95	0.999999985	0.61077553	0.954
y96	0.999999985	0.603709166	0.850
y97	0.999999985	0.543526792	0.971
y98	0.999999985	0.870710287	0.976
y99	0.999999985	0.917062944	0.944
y100	0.999999985	0.747657491	0.929
y101	0.999999985	0.813534793	0.872
y102	0.999999985	0.07535443	0.742
y103	0.999999985	0.953872896	0.776
y104	0.999999985	0.874444332	0.973
y105	0.999999985	0.904050843	0.905
y106	0.999999985	0.812269484	0.852
y107	0.999999985	0.797846925	0.834
y108	0.999999985	0.090498031	0.987
y109	0.999999985	0.649092702	0.946
y110	0.999999985	0.999568861	0.668
y111	0.999999985	0.998509508	0.939
y112	0.999999985	0.639970444	0.947
y113	0.999999985	0.796651362	0.938
y114	0.999999985	0.3992252	0.872
y115	0.999999985	0.398795989	0.871
y116	0.999999985	0.719225222	0.913
y117	0.999999985	0.623261406	0.859
y118	0.999999985	0.828466603	0.901

y119	0.999999985	0.715003707	0.904
y120	0.999999985	0.514307251	0.893
y121	0.999999985	0.85009573	0.939
y122	0.999999985	0.953707894	0.960
y123	0.999999985	0.821119365	0.941
y124	0.999999985	0.804824597	0.686
y125	0.999999985	0.687848828	0.912
y126	0.999999985	0.980980482	0.908
y127	0.999999985	0.9784122	0.782
y128	0.999999985	0.106661458	0.933
y129	0.999999985	0.894439549	0.833
y130	0.999999985	0.808424936	0.862
y131	0.999999985	0.892095324	0.804
y132	0.999999985	0.223472733	0.919
y133	0.999999985	0.324647319	0.804
y134	0.999999985	0.745398823	0.963
y135	0.999999985	0.693700221	0.790
y136	0.999999985	0.00836216	0.826
y137	0.999999985	0.274574629	0.848
y138	0.999999985	0.016794419	0.966
y139	0.999999985	0.63097059	0.931
y140	0.999999985	0.420103993	0.870
y141	0.999999985	0.540288507	0.838
y142	0.999999985	0.883287033	0.941
y143	0.999999985	0.972195314	0.854
y144	0.999999985	0.981199828	0.979
y145	0.999999985	0.246839014	0.852
y146	0.999999985	0.982456817	0.855
y147	0.999999985	0.745442276	0.800
y148	0.999999985	0.833961102	0.919
y149	0.999999985	0.951111472	0.910
y150	0.999999985	0.639446953	0.885
y151	0.999999985	0.936237525	0.909
y152	0.999999985	0.857457556	0.824
y153	0.999999985	0.804467074	0.967
y154	0.999999985	0.88549172	0.976
y155	0.999999985	0.881577028	0.964
y156	0.999999985	0.043079466	0.757
y157	0.999999985	0.499832251	0.907
y158	0.999999985	0.903755229	0.931
y159	0.999999985	0.365152487	0.818
y160	0.999999985	0.611941604	0.880

y161	0.999999985	0.593012205	0.998
y162	0.999999985	0.578451034	0.962
y163	0.999999985	0.578047653	0.954
y164	0.999999985	0.828368724	0.765
y165	0.999999985	0.802048576	0.863
y166	0.999999985	0.043618479	0.983
y167	0.999999985	0.45883582	0.877
y168	0.999999985	0.000714083	0.787
y169	0.999999985	0.339184371	0.830
y170	0.999999985	0.848020456	0.921
y171	0.999999985	0.969596637	0.792
y172	0.999999985	0.484499122	0.937
y173	0.999999985	0.482408161	0.862
y174	0.999999985	0.916739632	0.764
y175	0.999999985	0.441426527	0.964
y176	0.999999985	0.761370956	0.815
y177	0.999999985	0.887652052	0.929
y178	0.999999985	0.402837537	0.920
y179	0.999999985	0.867521426	0.810
y180	0.999999985	0.717408561	0.917
y181	0.999999985	0.973653031	0.850
y182	0.999999985	0.165894857	0.775
y183	0.999999985	0.723189659	0.679
y184	0.999999985	0.879650979	0.900
y185	0.999999985	0.741009875	0.980
y186	0.999999985	0.72703747	0.728
y187	0.999999985	0.893566879	0.715
y188	0.999999985	0.815455144	0.782
y189	0.999999985	0.965468668	0.964
y190	0.999999985	0.284853588	0.780
y191	0.999999985	0.812277643	0.820
y192	0.999999985	0.116461483	0.877
y193	0.999999985	0.232028004	0.916
y194	0.999999985	0.601643567	0.930
y195	0.999999985	0.931329463	0.976
y196	0.999999985	0.687023243	0.896
y197	0.999999985	0.994188819	0.901
y198	0.999999985	0.902589206	0.895
y199	0.999999985	0.777659375	0.956
y200	0.999999985	0.505780448	0.897
y201	0.999999985	0.768468047	0.807
y202	0.999999985	0.693831798	0.884

y203	0.999999985	0.60344285	0.956
y204	0.999999985	0.992954229	0.909
y205	0.999999985	0.097513685	0.656
y206	0.999999985	0.454230472	0.846
y207	0.999999985	0.785156256	0.872
y208	0.999999985	0.972885192	0.927
y209	0.999999985	0.894440656	0.951
y210	0.999999985	0.699367115	0.930
y211	0.999999985	0.781079839	0.894
y212	0.999999985	0.319303306	0.882
y213	0.999999985	0.475479135	0.836
y214	0.999999985	0.762253064	0.827
y215	0.999999985	0.921298423	0.923
y216	0.999999985	0.204756749	0.917
y217	0.999999985	0.31758042	0.955
y218	0.999999985	0.85412332	0.794
y219	0.999999985	0.884153741	0.757
y220	0.999999985	0.706742899	0.913
y221	0.999999985	0.931820119	0.964
y222	0.999999985	0.10212798	0.781
y223	0.999999985	0.84600139	0.953
y224	0.999999985	0.836473424	0.791
y225	0.999999985	0.928687676	0.926
y226	0.999999985	0.91580721	0.766
y227	0.999999985	0.920446141	0.904
y228	0.999999985	0.782886899	0.901
y229	0.999999985	0.99658187	0.925
y230	0.999999985	0.686315629	0.978
y231	0.999999985	0.956796024	0.938
y232	0.999999985	0.403574524	0.933
y233	0.999999985	0.874468374	0.868
y234	0.999999985	0.799508364	0.967
y235	0.999999985	0.902132383	0.865
y236	0.999999985	0.983710777	0.903
y237	0.999999985	0.619749594	0.911
y238	0.999999985	0.815454906	0.827
y239	0.999999985	0.690017143	0.684
y240	0.999999985	0.901971689	0.947
y241	0.999999985	0.96784284	0.859
y242	0.999999985	0.55561266	0.922
y243	0.999999985	0.124040307	0.913
y244	0.999999985	0.949279807	0.786

y245	0.999999985	0.003127678	0.890
y246	0.999999985	0.892140295	0.932
y247	0.999999985	0.280161128	0.961
y248	0.999999985	0.744465481	0.790
y249	0.999999985	0.813468024	0.991
y250	0.999999985	0.939791971	0.921
y251	0.999999985	0.868566294	0.897
y252	0.999999985	0.555707443	0.977
y253	0.999999985	0.512725586	0.674
y254	0.999999985	0.739132719	0.817
y255	0.999999985	0.841359763	0.890
y256	0.999999985	0.814583928	0.950
y257	0.999999985	0.897677065	0.973
y258	0.999999985	0.725957888	0.782
y259	0.999999985	0.790320314	0.982
y260	0.999999985	0.911559149	0.939
y261	0.999999985	0.606052389	0.804
y262	0.999999985	0.665655833	0.924
y263	0.999999985	0.571666685	0.861
y264	0.999999985	0.783104157	0.919
y265	0.999999985	0.584688199	0.829
y266	0.999999985	0.60886028	0.816
y267	0.999999985	0.815004004	0.915
y268	0.999999985	0.993484141	0.984
y269	0.999999985	0.77863733	0.975
y270	0.999999985	0.977010281	0.876
y271	0.999999985	0.904364411	0.879
y272	0.999999985	0.676053754	0.731
y273	0.999999985	0.879559936	0.902
y274	0.999999985	0.882969112	0.864
y275	0.999999985	0.802151145	0.985
y276	0.999999985	0.697735227	0.756
y277	0.999999985	0.917490588	0.922
y278	0.999999985	0.751355816	0.715
y279	0.999999985	0.837037109	0.861
y280	0.999999985	0.97141424	0.843
y281	0.999999985	0.914333102	0.990
y282	0.999999985	0.00226949	0.912
y283	0.999999985	0.83865296	0.895
y284	0.999999985	0.980729758	0.986
y285	0.999999985	0.835784157	0.715
y286	0.999999985	0.78212927	0.819

y287	0.999999985	6.92E-08	0.803
y288	0.999999985	0.894316001	0.989
y289	0.999999985	0.964701354	0.809
y290	0.999999985	0.392035769	0.941
y291	0.999999985	0.495333594	0.917
y292	0.999999985	0.888157433	0.994
y293	0.999999985	0.491001016	0.883
y294	0.999999985	0.885150184	0.977
y295	0.999999985	0.720950519	0.715
y296	0.999999985	0.703482236	0.946
y297	0.999999985	0.308482111	0.989
y298	0.999999985	0.729842661	0.786
y299	0.999999985	0.996516895	0.937
y300	0.999999985	0.67569961	0.970
y301	0.999999985	0.933928506	0.906
y302	0.999999985	0.825827937	0.940
y303	0.999999985	0.983358757	0.730
y304	0.999999985	0.660972859	0.933
y305	0.999999985	0.607405238	0.797
y306	0.999999985	0.986411542	0.936
y307	0.999999985	0.970758911	0.941
y308	0.999999985	0.974177516	0.901
y309	0.999999985	0.918090702	0.913
y310	0.999999985	0.717805097	0.855
y311	0.999999985	0.982308528	0.937
y312	0.999999985	0.500378222	0.813
y313	0.999999985	0.829211886	0.980
y314	0.999999985	0.991721666	0.967
y315	0.999999985	0.79309614	0.894
y316	0.999999985	0.828221371	0.647
y317	0.999999985	0.860825249	0.858
y318	0.999999985	0.942005062	0.942
y319	0.999999985	0.911686834	0.985
y320	0.999999985	0.958305313	0.997
y321	0.999999985	0.305583301	0.914
y322	0.999999985	0.967626324	0.723
y323	0.999999985	0.783235569	0.941
y324	0.999999985	0.924765707	0.846
y325	0.999999985	0.906551214	0.964
y326	0.999999985	0.973947142	0.866
y327	0.999999985	0.190070482	0.849
y328	0.999999985	0.627005713	0.895

y329	0.999999985	0.871678726	0.978
y330	0.999999985	0.981200228	0.982
y331	0.999999985	0.994233682	0.885
y332	0.999999985	0.870907518	0.865
y333	0.999999985	0.701508749	0.840
y334	0.999999985	0.876222887	0.989
y335	0.999999985	0.826151083	0.928
y336	0.999999985	0.678350623	0.910
y337	0.999999985	0.476510013	0.768
y338	0.999999985	0.652592276	0.732
y339	0.999999985	0.979014035	0.906
y340	0.999999985	0.866203364	0.850
y341	0.999999985	0.988487726	0.963
y342	0.999999985	0.975101657	0.998
y343	0.999999985	0.752191486	0.672
y344	0.999999985	0.878094528	0.744
y345	0.999999985	0.599012111	0.882
y346	0.999999985	0.829348962	0.845
y347	0.999999985	0.786430104	0.938
y348	0.999999985	0.858243908	0.955
y349	0.999999985	0.93837301	0.841
y350	0.999999985	0.956990321	0.992
y351	0.999999985	0.956025386	0.813
y352	0.999999985	0.812019053	0.957
y353	0.999999985	0.72340461	0.999
y354	0.999999985	0.961427015	0.768
y355	0.999999985	0.997908491	0.994
y356	0.999999985	0.898099464	0.923
y357	0.999999985	0.203942398	0.767
y358	0.999999985	0.489431373	0.955
y359	0.999999985	0.644276057	0.949
y360	0.999999985	0.896017201	0.784
y361	0.999999985	0.975106746	0.970
y362	0.999999985	0.974599255	0.844
y363	0.999999985	0.922655646	0.945
y364	0.999999985	0.962026041	0.980
y365	0.999999985	0.62589146	0.638
y366	0.999999985	0.947710532	0.936
y367	0.999999985	0.440366073	0.985
y368	0.999999985	0.769467387	0.947
y369	0.999999985	0.756169703	0.836
y370	0.999999985	0.883134383	0.889

y371	0.999999985	0.693371095	0.924
y372	0.999999985	0.836771632	0.967
y373	0.999999985	0.756964821	0.904
y374	0.999999985	0.82982928	0.805
y375	0.999999985	0.957597157	0.909
y376	0.999999985	0.888842249	0.956
y377	0.999999985	0.991096365	0.837
y378	0.999999985	0.799717914	0.749
y379	0.999999985	0.989607228	0.961
y380	0.999999985	0.8963122	0.816
y381	0.999999985	0.43030966	0.981
y382	0.999999985	0.748638201	0.922
y383	0.999999985	0.715309703	0.977
y384	0.999999985	0.880583055	0.912
y385	0.999999985	0.954906919	0.800
y386	0.999999985	0.973692331	0.999
y387	0.999999985	0.814377741	0.731
y388	0.999999985	0.818134988	0.711
y389	0.999999985	0.688685777	0.809
y390	0.999999985	0.893207768	0.839
y391	0.999999985	0.850063108	0.931
y392	0.999999985	0.87628495	0.871
y393	0.999999985	0.842841773	0.792
y394	0.999999985	0.853376889	0.811
y395	0.999999985	0.449321856	0.850
y396	0.999999985	0.877456846	0.939
y397	0.999999985	0.871559542	0.882
y398	0.999999985	0.686850499	0.922
y399	0.999999985	0.817959537	0.805
y400	0.999999985	0.881435296	0.821
y401	0.999999985	0.847718874	0.904
y402	0.999999985	0.991846667	0.958
y403	0.999999985	0.94660447	0.999
y404	0.999999985	0.905403911	0.932
y405	0.999999985	0.950327277	0.953
y406	0.999999985	0.90667855	0.939
y407	0.999999985	0.777309803	0.864
y408	0.999999985	0.910234916	0.952
y409	0.999999985	0.598351633	0.986
y410	0.999999985	0.78276153	0.831
y411	0.999999985	0.89349957	0.950
y412	0.999999985	0.784145216	0.855

y413	0.999999985	0.942105966	0.822
y414	0.999999985	0.792293875	0.744
y415	0.999999985	0.901923459	0.991
y416	0.999999985	0.954207362	0.985
y417	0.999999985	0.929056132	0.801
y418	0.999999985	0.929530884	0.829
y419	0.999999985	0.83760109	0.862
y420	0.999999985	0.906208796	0.832
y421	0.999999985	0.822857786	0.960
y422	0.999999985	0.8593319	0.833
y423	0.999999985	0.994662153	0.968
y424	0.999999985	0.740617433	0.815
y425	0.999999985	0.819736854	0.912
y426	0.999999985	0.99407927	0.604
y427	0.999999985	0.985341131	0.971
y428	0.999999985	0.889927803	0.928
y429	0.999999985	0.936696751	0.970
y430	0.999999985	0.938913702	0.795
y431	0.999999985	0.728991811	0.691
y432	0.999999985	0.604882305	0.837
y433	0.999999985	0.619982804	0.724
y434	0.999999985	0.586028913	0.724
y435	0.999999985	0.629646187	0.730
y436	0.999999985	0.913720068	0.979
y437	0.999999985	0.9738023	0.857
y438	0.999999985	0.955575963	0.786
y439	0.999999985	0.915019473	0.918
y440	0.999999985	0.85852196	0.881
y441	0.999999985	0.749239768	0.994
y442	0.999999985	0.88848126	0.948
y443	0.999999985	0.666925271	0.794
y444	0.999999985	0.766986863	0.912
y445	0.999999985	0.881709086	0.965
y446	0.999999985	0.775236332	0.887
y447	0.999999985	0.935219921	0.916
y448	0.999999985	0.953608993	0.920
y449	0.999999985	0.978125959	0.996
y450	0.999999985	0.844771476	0.830
y451	0.999999985	0.897851407	0.955
y452	0.999999985	0.983732084	0.950
y453	0.999999985	0.937480318	0.712
y454	0.999999985	0.981215259	0.734

y455	0.999999985	0.934554563	0.916
y456	0.999999985	0.915609627	0.808
y457	0.999999985	0.917677329	0.895
y458	0.999999985	0.809651243	0.778
y459	0.999999985	0.74002356	0.848
y460	0.999999985	0.667991648	0.915
y461	0.999999985	0.597227918	0.949
y462	0.999999985	0.832951417	0.801
y463	0.999999985	0.926376338	0.802
y464	0.999999985	0.936802864	0.910
y465	0.999999985	0.894564308	0.901
y466	0.999999985	0.959790262	0.991
y467	0.999999985	0.700763508	0.825
y468	0.999999985	0.946724641	0.991
y469	0.999999985	0.827642665	0.806
y470	0.999999985	0.766426546	0.841
y471	0.999999985	0.691333266	0.943
y472	0.999999985	0.891854896	0.946
y473	0.999999985	0.732122427	0.898
y474	0.999999985	0.907067447	0.903
y475	0.999999985	0.769467647	0.864
y476	0.999999985	0.892920025	0.980
y477	0.999999985	0.953177875	0.980
y478	0.999999985	0.856375929	0.989
y479	0.999999985	0.82861851	0.904
y480	0.999999985	0.824076611	0.991
y481	0.999999985	0.906870144	0.964
y482	0.999999985	0.954393173	0.987
y483	0.999999985	0.809198256	0.884
y484	0.999999985	0.985482567	0.977
y485	0.999999985	0.932135631	0.957
y486	0.999999985	0.831971439	0.945
y487	0.999999985	0.931360342	0.916
y488	0.999999985	0.811202927	0.853
y489	0.999999985	0.833835539	0.975
y490	0.999999985	0.907676164	0.924
y491	0.999999985	0.911291104	0.998
y492	0.999999985	0.690214719	0.977
y493	0.999999985	0.89473239	0.804
y494	0.999999985	0.675947646	0.919
y495	0.999999985	0.998827868	0.929
y496	0.999999985	0.861704956	0.998

y497	0.999999985	1	0.809
y498	0.999999985	0.899644949	0.846
y499	0.999999985	0.806803074	0.921
y500	0.999999985	0.761429683	0.825
y501	0.999999985	0.906626923	0.997
y502	0.999999985	0.965644688	0.945
y503	0.999999985	0.973266025	0.889
y504	0.999999985	0.939446196	0.976
y505	0.999999985	0.900529973	0.725
y506	0.999999985	0.979523372	0.997
y507	0.999999985	0.792882814	0.792
y508	0.999999985	0.997658215	0.959
y509	0.999999985	0.847171749	0.881
y510	0.999999985	0.894680542	0.820
y511	0.999999985	0.937475701	0.948
y512	0.999999985	0.942264131	0.883
y513	0.999999985	0.991075965	0.875
y514	0.999999985	0.827976496	0.942
y515	0.999999985	0.870515939	0.873
y516	0.999999985	0.808646253	0.877
y517	0.999999985	0.879630538	0.947
y518	0.999999985	0.789992453	0.882
y519	0.999999985	0.95013883	0.846
y520	0.999999985	0.990726009	0.800
y521	0.999999985	0.93297616	0.892
y522	0.999999985	0.906778795	0.849
y523	0.999999985	0.959307932	0.994
y524	0.999999985	0.892505149	0.875
y525	0.999999985	0.965340646	0.973
y526	0.999999985	0.983038985	0.922
y527	0.999999985	0.869845388	0.751
y528	0.999999985	0.911073607	0.977
y529	0.999999985	0.845529356	0.845
y530	0.999999985	0.93175352	0.870
y531	0.999999985	0.933412133	0.990
y532	0.999999985	0.946142945	0.965
y533	0.999999985	0.820870869	0.843
y534	0.999999985	0.933572592	0.885
y535	0.999999985	0.97816254	0.933
y536	0.999999985	0.998691142	0.793
y537	0.999999985	0.698251278	0.762
y538	0.999999985	0.819839928	0.814

y539	0.999999985	0.955198206	0.974
y540	0.999999985	0.92658717	0.889
y541	0.999999985	0.799758136	0.858
y542	0.999999985	0.703904386	0.719
y543	0.999999985	0.754067377	0.932
y544	0.999999985	0.99583151	0.995
y545	0.999999985	0.808421426	0.950
y546	0.999999985	0.939385256	0.972
y547	0.999999985	0.948142874	0.887
y548	0.999999985	0.927350512	0.988
y549	0.999999985	0.868063267	0.984
y550	0.999999985	0.793782033	0.924
y551	0.999999985	0.890551813	0.960
y552	0.999999985	0.972448799	0.933
y553	0.999999985	0.911380366	0.847
y554	0.999999985	0.774677061	0.877
y555	0.999999985	0.931407565	0.988
y556	0.999999985	0.867354537	0.861
y557	0.999999985	0.66964595	0.788
y558	0.999999985	0.980285405	0.995
y559	0.999999985	0.747680616	0.822
y560	0.999999985	0.823493512	0.776
y561	0.999999985	0.959653254	0.992
y562	0.999999985	0.860879554	0.825
y563	0.999999985	0.979572851	0.928
y564	0.999999985	0.980628716	0.971
y565	0.999999985	0.819050891	0.823
y566	0.999999985	0.990354004	0.938
y567	0.999999985	0.96440238	0.927
y568	0.999999985	0.989881483	0.902
y569	0.999999985	0.871869824	0.974
y570	0.999999985	0.902312323	0.860
y571	0.999999985	0.830513876	0.740
y572	0.999999985	0.922015619	0.896
y573	0.999999985	0.843457421	0.774
y574	0.999999985	0.963016741	0.921
y575	0.999999985	0.909499771	0.904
y576	0.999999985	0.978531035	0.999
y577	0.999999985	0.946238203	0.982
y578	0.999999985	0.965795828	0.981
y579	0.999999985	0.953417217	0.963
y580	0.999999985	0.984574293	0.949

y581	0.999999985	0.881316544	0.785
y582	0.999999985	0.774665563	0.993
y583	0.999999985	0.812429311	0.708
y584	0.999999985	0.950900947	0.949
y585	0.999999985	0.800163612	0.747
y586	0.999999985	0.915925876	0.885
y587	0.999999985	0.816202726	0.855
y588	0.999999985	0.930975628	0.844
y589	0.999999985	0.928227998	0.954
y590	0.999999985	0.885393751	0.998
y591	0.999999985	0.807778457	0.700
y592	0.999999985	0.929055664	0.891

Appendix D Chapter 7: Taxa of the Saliva Microbiota

Supplementary table D 7.1 Taxa identified in the saliva microbiota of the cohort

	Family	Genus	Species
ASV1	Pasteurellaceae	Haemophilus	influenzae/parainfluenzae
ASV2	Neisseriaceae	Neisseria	cinerea/flavescens/lactamica/meningitidis/mucosa/perflava/subflava
ASV3	Prevotellaceae	Prevotella_7	melaninogenica
ASV4	Streptococcaceae	Streptococcus	anginosus/cristatus/gordonii/infantis/mitis/oralis/pneumoniae/pseudopneumoniae/sanguinis/timonensis
ASV5	Veillonellaceae	Veillonella	atypica/dispar
ASV6	Fusobacteriaceae	Fusobacterium	periodonticum
ASV7	Campylobacteraceae	Campylobacter	concisus
ASV8	Porphyromonadaceae	Porphyromonas	pasteri
ASV9	Streptococcaceae	Streptococcus	australis/mitis/oralis/parasanguinis/rubneri/sanguinis
ASV10	Prevotellaceae	Prevotella_7	histicola
ASV11	Pasteurellaceae	Aggregatibacter	segnis
ASV12	Prevotellaceae	Prevotella_6	salivae
ASV13	Prevotellaceae	Alloprevotella	NA
ASV14	Carnobacteriaceae	Granulicatella	adiacens/para-adiacens
ASV15	Family_XI	Gemella	haemolysans/morbillorum/parahaemolysans/sanguinis/taiwanensis
ASV16	Pasteurellaceae	Haemophilus	haemolyticus/influenzae/parainfluenzae
ASV17	Veillonellaceae	Veillonella	parvula/rogosae/tobetsuensis

ASV18	Streptococcaceae	Streptococcus	australis/infantis/lactarius/mitis/oralis/peroris/sanguinis
ASV19	Prevotellaceae	Prevotella	nanceiensis
ASV20	Neisseriaceae	Neisseria	flava/lactamica/macacae/meningitidis/mucosa/perflava/pharyngis/polysaccharea/sicca
ASV21	Micrococcaceae	Rothia	mucilaginosa
ASV22	Prevotellaceae	Prevotella_7	jejuni/melaninogenica
ASV23	Micrococcaceae	Rothia	mucilaginosa
ASV24	Leptotrichiaceae	Leptotrichia	NA
ASV25	Streptococcaceae	Streptococcus	NA
ASV26	Micrococcaceae	Rothia	NA
ASV27	Fusobacteriaceae	Fusobacterium	nucleatum
ASV28	Micrococcaceae	Rothia	mucilaginosa
ASV29	Pasteurellaceae	Actinobacillus	NA
ASV30	Prevotellaceae	Prevotella	pallens
ASV31	Veillonellaceae	Veillonella	atypica/parvula
ASV32	Prevotellaceae	Alloprevotella	NA
ASV33	Prevotellaceae	Prevotella	nanceiensis
ASV34	Actinomycetaceae	Actinomyces	odontolyticus
ASV35	Pasteurellaceae	Mannheimia	NA
ASV36	Veillonellaceae	Megasphaera	micronuciformis
ASV37	Prevotellaceae	Prevotella_7	NA
ASV38	Prevotellaceae	Prevotella	NA
ASV39	Fusobacteriaceae	Fusobacterium	NA
ASV40	Pasteurellaceae	Actinobacillus	porcinus
ASV41	Prevotellaceae	Alloprevotella	NA
ASV42	Veillonellaceae	Selenomonas_3	NA
ASV43	Micrococcaceae	Rothia	dentocariosa

ASV44	Neisseriaceae	Neisseria	flavescens
ASV45	Veillonellaceae	Veillonella	dispar/parvula
ASV46	Weeksellaceae	Bergeyella	NA
ASV47	Actinomycetaceae	Actinomyces	odontolyticus
ASV48	Fusobacteriaceae	Fusobacterium	canifelinum/nucleatum
ASV49	Neisseriaceae	Kingella	NA
ASV50	Actinomycetaceae	Actinomyces	graevenitzii
ASV51	Atopobiaceae	Atopobium	parvulum
ASV52	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV53	Leptotrichiaceae	Oceanivirga	NA
ASV54	Campylobacteraceae	Campylobacter	rectus/showae
ASV55	Neisseriaceae	Neisseria	flavescens
ASV56	Corynebacteriaceae	Corynebacterium	matruchotii
ASV57	Neisseriaceae	Neisseria	NA
ASV58	Lachnospiraceae	Oribacterium	sinus
ASV59	Leptotrichiaceae	Leptotrichia	NA
ASV60	Prevotellaceae	Alloprevotella	rava
ASV61	NA	NA	NA
ASV62	Lachnospiraceae	Stomatobaculum	NA
ASV63	Lactobacillaceae	Lactobacillus	delbrueckii/fermentum/ingluviei/oris/parabuchneri/plantarum
ASV64	Fusobacteriaceae	Fusobacterium	naviforme/nucleatum
ASV65	Burkholderiaceae	Lautropia	mirabilis
ASV66	Prevotellaceae	Alloprevotella	NA
ASV67	Flavobacteriaceae	Capnocytophaga	leadbetteri
ASV68	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV69	Veillonellaceae	Veillonella	NA

ASV70	Lachnospiraceae	Lachnoanaerobaculum	orale/saburreum
ASV71	Campylobacteraceae	Campylobacter	rectus/showae
ASV72	Enterobacteriaceae	Salmonella	NA
ASV73	Enterobacteriaceae	Escherichia/Shigella	NA
ASV74	Streptococcaceae	Streptococcus	cristatus/sanguinis
ASV75	Pasteurellaceae	Aggregatibacter	aphrophilus
ASV76	Lachnospiraceae	NA	NA
ASV77	Bacillaceae	Bacillus	altitudinis/amyloliquefaciens/firmus/halotolerans/licheniformis/mojavensis/siamensis/subtilis/tequilensis/timonensis/vallismortis/velezensis
ASV78	Fusobacteriaceae	Fusobacterium	nucleatum
ASV79	Leptotrichiaceae	Leptotrichia	NA
ASV80	Campylobacteraceae	Campylobacter	concisus
ASV81	Pasteurellaceae	Mannheimia	NA
ASV82	Flavobacteriaceae	Capnocytophaga	gingivalis
ASV83	Porphyromonadaceae	Porphyromonas	endodontalis
ASV84	Pseudomonadaceae	Pseudomonas	aeruginosa/denitrificans/fluorescens/indica/mendocina/otitidis/protegens/putida/tropicalis
ASV85	Erysipelotrichaceae	Solobacterium	moorei
ASV86	Prevotellaceae	Prevotella	oris
ASV87	Lachnospiraceae	Stomatobaculum	longum
ASV88	Prevotellaceae	Prevotella_6	NA
ASV89	Prevotellaceae	Prevotella	nigrescens
ASV90	Lachnospiraceae	Lachnoanaerobaculum	cf.
ASV91	Porphyromonadaceae	Porphyromonas	gingivalis
ASV92	Prevotellaceae	Alloprevotella	NA
ASV93	Prevotellaceae	Prevotella_2	shahii
ASV94	Campylobacteraceae	Campylobacter	gracilis
ASV95	Pasteurellaceae	Haemophilus	NA

ASV96	Paludibacteraceae	F0058	NA
ASV97	Actinomycetaceae	Actinomyces	naeslundii/oris/viscosus
ASV98	Prevotellaceae	Prevotella_7	denticola
ASV99	Fusobacteriaceae	Fusobacterium	nucleatum
ASV100	Micrococcaceae	Rothia	aeria/dentocariosa
ASV101	Spirochaetaceae	Treponema_2	denticola/putidum
ASV102	Flavobacteriaceae	Capnocytophaga	granulosa
ASV103	Burkholderiaceae	Lautropia	NA
ASV104	Campylobacteraceae	Campylobacter	NA
ASV105	Pasteurellaceae	Haemophilus	NA
ASV106	Prevotellaceae	Prevotella	NA
ASV107	Flavobacteriaceae	Capnocytophaga	sputigena
ASV108	Leptotrichiaceae	Leptotrichia	NA
ASV109	Prevotellaceae	Alloprevotella	rava
ASV110	Fusobacteriaceae	Fusobacterium	NA
ASV111	Veillonellaceae	Dialister	pneumosintes
ASV112	NA	NA	NA
ASV113	Prevotellaceae	Alloprevotella	tannerae
ASV114	Prevotellaceae	Prevotella_7	NA
ASV115	Lachnospiraceae	Catonella	NA
ASV116	Neisseriaceae	Neisseria	bacilliformis/lactamica/meningitidis
ASV117	Family_XIII	NA	NA
ASV118	Neisseriaceae	Neisseria	oralis
ASV119	Prevotellaceae	Prevotella	NA
ASV120	Veillonellaceae	Veillonella	NA
ASV121	Neisseriaceae	Kingella	oralis

ASV122	Actinomycetaceae	Actinomyces	NA
ASV123	NA	NA	NA
ASV124	Spirochaetaceae	Treponema_2	socranskii
ASV125	Weeksellaceae	Bergeyella	NA
ASV126	Neisseriaceae	Neisseria	oralis
ASV127	Lachnospiraceae	Oribacterium	asaccharolyticum
ASV128	Mycoplasmataceae	Mycoplasma	falconis/faucium/hyosynoviae
ASV129	Family_XIII	Mogibacterium	diversum/neglectum/pumilum/vescum
ASV130	Pasteurellaceae	Actinobacillus	NA
ASV131	Fusobacteriaceae	Fusobacterium	NA
ASV132	Prevotellaceae	Prevotella	NA
ASV133	Prevotellaceae	Prevotella	NA
ASV134	Carnobacteriaceae	Granulicatella	elegans
ASV135	Family_XI	Parvimonas	micra
ASV136	Veillonellaceae	Selenomonas	sputigena
ASV137	Campylobacteraceae	Campylobacter	gracilis
ASV138	Peptostreptococcaceae	Peptostreptococcus	stomatis
ASV139	Prevotellaceae	Prevotella	oulorum
ASV140	Veillonellaceae	Dialister	invisus
ASV141	Prevotellaceae	Prevotella	NA
ASV142	Lachnospiraceae	Oribacterium	parvum
ASV143	Corynebacteriaceae	Corynebacterium	durum
ASV144	Lactobacillaceae	Lactobacillus	crispatus/gasseri/helveticus/hominis/iatae/johnsonii/kefiranofaciens/prophage/taiwanensis
ASV145	Lentimicrobiaceae	NA	NA
ASV146	Prevotellaceae	Prevotella	NA
ASV147	Family_XI	Parvimonas	micra

ASV148	Prevotellaceae	Prevotella_2	conceptionensis
ASV149	Prevotellaceae	Prevotella_7	melaninogenica
ASV150	Pasteurellaceae	Haemophilus	NA
ASV151	Leptotrichiaceae	Leptotrichia	NA
ASV152	Prevotellaceae	Prevotella_7	baroniae
ASV153	Listeriaceae	Listeria	innocua/ivanovii/marthii/monocytogenes/phage/seeligeri/welshimeri
ASV154	Prevotellaceae	Prevotella	NA
ASV155	Veillonellaceae	Veillonella	denticariosi/parvula
ASV156	Neisseriaceae	Neisseria	NA
ASV157	Campylobacteraceae	Campylobacter	showae
ASV158	Veillonellaceae	Selenomonas_3	infelix
ASV159	Veillonellaceae	Veillonella	NA
ASV160	Aerococcaceae	Abiotrophia	defectiva
ASV161	Cardiobacteriaceae	Cardiobacterium	hominis
ASV162	NA	NA	NA
ASV163	Neisseriaceae	Neisseria	NA
ASV164	Leptotrichiaceae	Leptotrichia	wadei
ASV165	Prevotellaceae	Prevotella_7	NA
ASV166	Spirochaetaceae	Treponema_2	NA
ASV167	Prevotellaceae	Prevotella_7	veroralis
ASV168	Leptotrichiaceae	Leptotrichia	hongkongensis
ASV169	Spirochaetaceae	Treponema_2	lecithinolyticum
ASV170	Prevotellaceae	Prevotella_7	multiformis
ASV171	Neisseriaceae	Neisseria	mucosa/pharyngis
ASV172	NA	NA	NA
ASV173	Peptococcaceae	Peptococcus	NA

ASV174	Weeksellaceae	Bergeyella	NA
ASV175	Lachnospiraceae	Stomatobaculum	NA
ASV176	Enterobacteriaceae	NA	NA
ASV177	Lachnospiraceae	Butyrivibrio_2	NA
ASV178	Veillonellaceae	Veillonella	NA
ASV179	Pasteurellaceae	Haemophilus	haemolyticus/influenzae/quentini
ASV180	Veillonellaceae	Selenomonas_3	noxia
ASV181	Pasteurellaceae	Rodentibacter	NA
ASV182	Enterococcaceae	Enterococcus	NA
ASV183	Pasteurellaceae	Haemophilus	NA
ASV184	Veillonellaceae	Veillonella	NA
ASV185	Prevotellaceae	Prevotella	maculosa
ASV186	Leptotrichiaceae	Leptotrichia	shahii/wadei
ASV187	Porphyromonadaceae	Porphyromonas	catoniae
ASV188	Tannerellaceae	Tannerella	NA
ASV189	Tannerellaceae	Tannerella	forsythia
ASV190	Spirochaetaceae	Treponema_2	amylovorum
ASV191	Spirochaetaceae	Treponema_2	NA
ASV192	Prevotellaceae	Prevotella	intermedia
ASV193	Pasteurellaceae	Haemophilus	NA
ASV194	NA	NA	NA
ASV195	Fusobacteriaceae	Fusobacterium	NA
ASV196	Neisseriaceae	Neisseria	NA
ASV197	Prevotellaceae	Prevotella_7	dentalis
ASV198	Prevotellaceae	Prevotella	NA
ASV199	Actinomycetaceae	F0332	NA

ASV200	Tannerellaceae	Tannerella	NA
ASV201	Weeksellaceae	Bergeyella	NA
ASV202	Neisseriaceae	Alysiella	NA
ASV203	Prevotellaceae	Prevotella	NA
ASV204	Veillonellaceae	NA	NA
ASV205	Mycoplasmataceae	Mycoplasma	salivarium
ASV206	Pasteurellaceae	Rodentibacter	trehalosifermentans
ASV207	Campylobacteraceae	Campylobacter	NA
ASV208	Carnobacteriaceae	Granulicatella	NA
ASV209	Prevotellaceae	Prevotella_7	NA
ASV210	Weeksellaceae	Bergeyella	NA
ASV211	Staphylococcaceae	Staphylococcus	argenteus/aureus/capitis/caprae/devriesei/epidermidis/equorum/haemolyticus/hominis/lugdunensis/pasteuri/petrasii/phage/saccharolyticus
ASV212	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV213	Streptococcaceae	Streptococcus	NA
ASV214	Veillonellaceae	Veillonella	NA
ASV215	Synergistaceae	Fretibacterium	fastidiosum
ASV216	Prevotellaceae	Alloprevotella	tannerae
ASV217	Leptotrichiaceae	Leptotrichia	buccalis
ASV218	Fusobacteriaceae	Fusobacterium	NA
ASV219	Prevotellaceae	Prevotella_7	NA
ASV220	Neisseriaceae	Neisseria	NA
ASV221	Veillonellaceae	Selenomonas_3	NA
ASV222	Prevotellaceae	NA	NA
ASV223	NA	NA	NA
ASV224	Neisseriaceae	Neisseria	meningitidis
ASV225	Spirochaetaceae	Treponema_2	vincentii

ASV226	Veillonellaceae	Selenomonas_3	flueggei/flueggei-like
ASV227	Prevotellaceae	Prevotella_7	NA
ASV228	Porphyromonadaceae	Porphyromonas	NA
ASV229	Burkholderiaceae	Comamonas	NA
ASV230	Porphyromonadaceae	Porphyromonas	gingivalis
ASV231	Leptotrichiaceae	Leptotrichia	NA
ASV232	Prevotellaceae	Alloprevotella	NA
ASV233	Neisseriaceae	Neisseria	NA
ASV234	Flavobacteriaceae	Capnocytophaga	NA
ASV235	Pasteurellaceae	Haemophilus	NA
ASV236	Neisseriaceae	Eikenella	corrodens
ASV237	Spirochaetaceae	Treponema_2	medium
ASV238	Lachnospiraceae	Catonella	morbi
ASV239	Cardiobacteriaceae	Cardiobacterium	valvarum
ASV240	Spirochaetaceae	Treponema_2	NA
ASV241	Neisseriaceae	NA	NA
ASV242	Lachnospiraceae	Oribacterium	NA
ASV243	Lachnospiraceae	Johnsonella	NA
ASV244	Synergistaceae	Fretibacterium	NA
ASV245	NA	NA	NA
ASV246	Spirochaetaceae	Treponema_2	NA
ASV247	Streptococcaceae	Streptococcus	anginosus/constellatus/intermedius
ASV248	Prevotellaceae	Prevotella_7	NA
ASV249	Pasteurellaceae	Haemophilus	NA
ASV250	Veillonellaceae	Selenomonas_3	artemidis
ASV251	Prevotellaceae	Alloprevotella	NA

ASV252	Veillonellaceae	Selenomonas	NA
ASV253	Leptotrichiaceae	Leptotrichia	NA
ASV254	Atopobiaceae	Atopobium	NA
ASV255	Prevotellaceae	Prevotella_7	NA
ASV256	Leptotrichiaceae	Leptotrichia	NA
ASV257	Prevotellaceae	Prevotella_7	NA
ASV258	Leptotrichiaceae	Leptotrichia	NA
ASV259	Flavobacteriaceae	Capnocytophaga	NA
ASV260	Streptococcaceae	Streptococcus	anginosus/constellatus/intermedius/lutetiensis
ASV261	Prevotellaceae	Prevotella_2	loescheii
ASV262	Leptotrichiaceae	Leptotrichia	hofstadii
ASV263	Saccharimonadaceae	NA	NA
ASV264	Prevotellaceae	Prevotella	intermedia
ASV265	Leptotrichiaceae	Leptotrichia	NA
ASV266	Streptococcaceae	Streptococcus	anginosus/constellatus/intermedius/milleri/mitis/sanguinis/suis
ASV267	Prevotellaceae	Prevotella_7	NA
ASV268	Family_XIII	NA	NA
ASV269	Campylobacteraceae	Campylobacter	showae
ASV270	Streptococcaceae	Streptococcus	mutans
ASV271	Lentimicrobiaceae	NA	NA
ASV272	Porphyromonadaceae	Porphyromonas	NA
ASV273	Prevotellaceae	Prevotella	aurantiaca
ASV274	Prevotellaceae	Prevotella_7	NA
ASV275	Spirochaetaceae	Treponema_2	NA
ASV276	Campylobacteraceae	Campylobacter	NA
ASV277	Streptococcaceae	Streptococcus	mitis/phage/pneumoniae/pseudopneumoniae

ASV278	Clostridiales_vadinBB60_group	NA	NA
ASV279	Neisseriaceae	Kingella	NA
ASV280	Fusobacteriaceae	Fusobacterium	NA
ASV281	Fusobacteriaceae	Fusobacterium	NA
ASV282	Prevotellaceae	Prevotella_2	NA
ASV283	Prevotellaceae	Prevotella_2	saccharolytica
ASV284	Fusobacteriaceae	Fusobacterium	nucleatum
ASV285	Prevotellaceae	Prevotella_7	NA
ASV286	Pasteurellaceae	Haemophilus	NA
ASV287	Spirochaetaceae	Treponema_2	NA
ASV288	Prevotellaceae	Prevotella_7	NA
ASV289	Spirochaetaceae	Treponema_2	NA
ASV290	Prevotellaceae	Prevotella_7	NA
ASV291	Pasteurellaceae	Aggregatibacter	NA
ASV292	Prevotellaceae	Prevotella	NA
ASV293	Pasteurellaceae	Haemophilus	NA
ASV294	Porphyromonadaceae	Porphyromonas	NA
ASV295	Prevotellaceae	Alloprevotella	NA
ASV296	Veillonellaceae	Anaeroglobus	geminatus
ASV297	Pasteurellaceae	Aggregatibacter	NA
ASV298	Neisseriaceae	Eikenella	NA
ASV299	Mycoplasmataceae	Mycoplasma	indiense/orale
ASV300	Lachnospiraceae	Oribacterium	NA
ASV301	Flavobacteriaceae	Capnocytophaga	NA
ASV302	Actinomycetaceae	Actinomyces	johnsonii/naeslundii/oris
ASV303	Peptostreptococcaceae	Filifactor	alocis

ASV304	Peptostreptococcaceae	Peptoanaerobacter	NA
ASV305	Fusobacteriaceae	Fusobacterium	massiliense
ASV306	Actinomycetaceae	Actinomyces	NA
ASV307	Leptotrichiaceae	Leptotrichia	NA
ASV308	Porphyromonadaceae	Porphyromonas	NA
ASV309	Flavobacteriaceae	Capnocytophaga	granulosa
ASV310	Flavobacteriaceae	Capnocytophaga	NA
ASV311	Pasteurellaceae	Haemophilus	NA
ASV312	Porphyromonadaceae	Porphyromonas	NA
ASV313	Fusobacteriaceae	Fusobacterium	NA
ASV314	Spirochaetaceae	Treponema_2	NA
ASV315	Prevotellaceae	Prevotella_7	NA
ASV316	Pasteurellaceae	Aggregatibacter	actinomycetemcomitans
ASV317	Veillonellaceae	Veillonella	NA
ASV318	Flavobacteriaceae	Capnocytophaga	NA
ASV319	Prevotellaceae	Alloprevotella	NA
ASV320	Bifidobacteriaceae	Scardovia	wiggsiae
ASV321	Spirochaetaceae	Treponema_2	maltophilum
ASV322	Leptotrichiaceae	Leptotrichia	NA
ASV323	Prevotellaceae	Prevotella_7	NA
ASV324	Synergistaceae	Fretibacterium	NA
ASV325	Family_XIII	NA	NA
ASV326	Porphyromonadaceae	Porphyromonas	NA
ASV327	Leptotrichiaceae	Leptotrichia	NA
ASV328	Campylobacteraceae	Campylobacter	NA
ASV329	Pasteurellaceae	Haemophilus	NA

ASV330	Streptococcaceae	Streptococcus	NA
ASV331	Lachnospiraceae	Oribacterium	NA
ASV332	Flavobacteriaceae	Capnocytophaga	ochracea
ASV333	Fusobacteriaceae	Fusobacterium	NA
ASV334	Prevotellaceae	Prevotella	oralis/phocaeensis
ASV335	Bifidobacteriaceae	Parascardovia	denticolens
ASV336	Streptococcaceae	Streptococcus	NA
ASV337	Lachnospiraceae	Catonella	NA
ASV338	Campylobacteraceae	Campylobacter	NA
ASV339	Spirochaetaceae	Treponema_2	medium
ASV340	Spirochaetaceae	Treponema_2	NA
ASV341	Spirochaetaceae	Treponema_2	NA
ASV342	Fusobacteriaceae	Fusobacterium	NA
ASV343	Spirochaetaceae	Treponema_2	vincentii
ASV344	Prevotellaceae	Prevotella	NA
ASV345	Spirochaetaceae	Treponema_2	medium/pallidum
ASV346	Micrococcaceae	Rothia	NA
ASV347	Leptotrichiaceae	Leptotrichia	NA
ASV348	Paludibacteraceae	F0058	NA
ASV349	Spirochaetaceae	Treponema_2	socranskii
ASV350	Prevotellaceae	Prevotella_7	NA
ASV351	Pasteurellaceae	Haemophilus	NA
ASV352	Leptotrichiaceae	Leptotrichia	NA
ASV353	Pasteurellaceae	Haemophilus	NA
ASV354	Fusobacteriaceae	Fusobacterium	NA
ASV355	Weeksellaceae	Bergeyella	NA

ASV356	Prevotellaceae	Prevotella	NA
ASV357	Flavobacteriaceae	Capnocytophaga	NA
ASV358	Spirochaetaceae	Treponema_2	NA
ASV359	Flavobacteriaceae	Capnocytophaga	gingivalis
ASV360	Streptococcaceae	Streptococcus	NA
ASV361	Pasteurellaceae	Haemophilus	haemolyticus
ASV362	Prevotellaceae	Prevotella_7	NA
ASV363	Prevotellaceae	Alloprevotella	NA
ASV364	Lachnospiraceae	Catonella	NA
ASV365	Lachnospiraceae	Catonella	NA
ASV366	Weeksellaceae	Bergeyella	NA
ASV367	Enterobacteriaceae	Klebsiella	NA
ASV368	Prevotellaceae	Alloprevotella	NA
ASV369	Acholeplasmataceae	Acholeplasma	NA
ASV370	Pasteurellaceae	Actinobacillus	NA
ASV371	Weeksellaceae	Bergeyella	NA
ASV372	Veillonellaceae	Veillonella	NA
ASV373	Neisseriaceae	Conchiformibius	NA
ASV374	Micrococcaceae	Rothia	mucilaginosa
ASV375	Veillonellaceae	Veillonella	NA
ASV376	Prevotellaceae	Alloprevotella	NA
ASV377	Bifidobacteriaceae	Alloscardovia	omnicolens
ASV378	Spirochaetaceae	Treponema_2	NA
ASV379	Spirochaetaceae	Treponema_2	NA
ASV380	Bacteroidales_Incertae_Sedis	Phocaeicola	abscessus
ASV381	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA

ASV382	Veillonellaceae	Veillonella	NA
ASV383	Actinomycetaceae	Actinomyces	NA
ASV384	Prevotellaceae	Alloprevotella	NA
ASV385	Prevotellaceae	Prevotella_7	NA
ASV386	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV387	Pasteurellaceae	Haemophilus	NA
ASV388	Fusobacteriaceae	Fusobacterium	NA
ASV389	Pasteurellaceae	Haemophilus	NA
ASV390	Prevotellaceae	Alloprevotella	NA
ASV391	Leptotrichiaceae	Leptotrichia	NA
ASV392	Veillonellaceae	Centipeda	NA
ASV393	Flavobacteriaceae	Capnocytophaga	haemolytica
ASV394	Prevotellaceae	Prevotella	NA
ASV395	Pasteurellaceae	Haemophilus	NA
ASV396	Fusobacteriaceae	Fusobacterium	NA
ASV397	Spirochaetaceae	Treponema_2	maltophilum
ASV398	Prevotellaceae	Alloprevotella	NA
ASV399	Neisseriaceae	NA	NA
ASV400	Pasteurellaceae	Haemophilus	NA
ASV401	Prevotellaceae	Prevotella	NA
ASV402	Veillonellaceae	Anaeroglobus	NA
ASV403	Prevotellaceae	Prevotella	NA
ASV404	Prevotellaceae	Prevotella_2	saccharolytica
ASV405	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV406	Veillonellaceae	Anaeroglobus	NA
ASV407	Fusobacteriaceae	Fusobacterium	NA

ASV408	Prevotellaceae	Prevotella	NA
ASV409	Neisseriaceae	Neisseria	NA
ASV410	Carnobacteriaceae	Granulicatella	NA
ASV411	Pasteurellaceae	Aggregatibacter	NA
ASV412	Spirochaetaceae	Treponema_2	NA
ASV413	Peptostreptococcaceae	NA	NA
ASV414	Prevotellaceae	Prevotella_2	marshii
ASV415	Weeksellaceae	Bergeyella	NA
ASV416	Prevotellaceae	Alloprevotella	NA
ASV417	Atopobiaceae	Atopobium	rimae
ASV418	Erysipelotrichaceae	Bulleidia	extructa
ASV419	Prevotellaceae	Prevotella_7	NA
ASV420	Leptotrichiaceae	Leptotrichia	goodfellowii
ASV421	Veillonellaceae	Selenomonas_3	NA
ASV422	Pasteurellaceae	Haemophilus	NA
ASV423	Veillonellaceae	Veillonella	NA
ASV424	Neisseriaceae	Neisseria	NA
ASV425	Fusobacteriaceae	Fusobacterium	NA
ASV426	Prevotellaceae	Prevotella	micans
ASV427	Flavobacteriaceae	Capnocytophaga	ochracea
ASV428	Spirochaetaceae	Treponema_2	NA
ASV429	Prevotellaceae	Prevotella_7	NA
ASV430	Pasteurellaceae	Haemophilus	NA
ASV431	Pasteurellaceae	Haemophilus	NA
ASV432	Prevotellaceae	Prevotella_7	NA
ASV433	Neisseriaceae	Neisseria	NA

ASV434	Neisseriaceae	Neisseria	NA
ASV435	Campylobacteraceae	Campylobacter	NA
ASV436	Campylobacteraceae	Campylobacter	curvus
ASV437	Weeksellaceae	Bergeyella	NA
ASV438	Prevotellaceae	Prevotella_7	NA
ASV439	Prevotellaceae	Prevotella_2	NA
ASV440	Pasteurellaceae	Mannheimia	NA
ASV441	Pasteurellaceae	Haemophilus	haemolyticus/influenzae
ASV442	Lentimicrobiaceae	NA	NA
ASV443	Leptotrichiaceae	Leptotrichia	NA
ASV444	Prevotellaceae	Alloprevotella	NA
ASV445	Leptotrichiaceae	Leptotrichia	NA
ASV446	Spirochaetaceae	Treponema_2	NA
ASV447	Streptococcaceae	Streptococcus	NA
ASV448	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV449	Lachnospiraceae	Oribacterium	NA
ASV450	Prevotellaceae	NA	NA
ASV451	Pasteurellaceae	Haemophilus	NA
ASV452	Veillonellaceae	Veillonella	NA
ASV453	Leptotrichiaceae	Leptotrichia	NA
ASV454	Prevotellaceae	Prevotella_7	NA
ASV455	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV456	Fusobacteriaceae	Fusobacterium	NA
ASV457	Neisseriaceae	Neisseria	NA
ASV458	Prevotellaceae	Prevotella_6	NA
ASV459	Streptococcaceae	Streptococcus	NA

ASV460	Neisseriaceae	Neisseria	NA
ASV461	Fusobacteriaceae	Fusobacterium	NA
ASV462	Veillonellaceae	Selenomonas_4	NA
ASV463	Leptotrichiaceae	Streptobacillus	NA
ASV464	Peptostreptococcaceae	Peptoanaerobacter	stomatis
ASV465	Prevotellaceae	Alloprevotella	NA
ASV466	Leptotrichiaceae	Leptotrichia	NA
ASV467	Fusobacteriaceae	Fusobacterium	NA
ASV468	NA	NA	NA
ASV469	Veillonellaceae	Veillonella	NA
ASV470	Fusobacteriaceae	Fusobacterium	necrophorum
ASV471	Fusobacteriaceae	Fusobacterium	NA
ASV472	Prevotellaceae	Prevotella_7	NA
ASV473	Neisseriaceae	Simonsiella	muelleri
ASV474	Prevotellaceae	Prevotella	NA
ASV475	Streptococcaceae	Streptococcus	NA
ASV476	Veillonellaceae	Veillonella	NA
ASV477	Neisseriaceae	Neisseria	NA
ASV478	Flavobacteriaceae	Capnocytophaga	NA
ASV479	Fusobacteriaceae	Fusobacterium	NA
ASV480	Prevotellaceae	Alloprevotella	NA
ASV481	Family_XIII	Mogibacterium	timidum
ASV482	Prevotellaceae	Prevotella_7	NA
ASV483	Flavobacteriaceae	Capnocytophaga	ochracea
ASV484	Neisseriaceae	Neisseria	NA
ASV485	Atopobiaceae	Atopobium	parvulum

ASV486	Pasteurellaceae	Aggregatibacter	NA
ASV487	Prevotellaceae	Prevotella_6	NA
ASV488	Weeksellaceae	Bergeyella	NA
ASV489	Lactobacillaceae	Lactobacillus	vaginalis
ASV490	Prevotellaceae	Prevotella_7	NA
ASV491	Carnobacteriaceae	Granulicatella	NA
ASV492	Prevotellaceae	Prevotella_7	NA
ASV493	Bifidobacteriaceae	Bifidobacterium	breve/kashiwanohense/longum/saguini
ASV494	Flavobacteriaceae	Capnocytophaga	NA
ASV495	Pasteurellaceae	Haemophilus	NA
ASV496	Spirochaetaceae	Treponema_2	NA
ASV497	Lactobacillaceae	Lactobacillus	casei/helveticus/paracasei/plantarum/rhamnosus/zeae
ASV498	Leptotrichiaceae	Oceanivirga	NA
ASV499	Prevotellaceae	Prevotella	NA
ASV500	Actinomycetaceae	F0332	NA
ASV501	Pasteurellaceae	Haemophilus	NA
ASV502	Prevotellaceae	Prevotella_7	NA
ASV503	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV504	Prevotellaceae	Prevotella_6	NA
ASV505	Pasteurellaceae	Haemophilus	aegyptius/influenzae/phage
ASV506	Pasteurellaceae	Haemophilus	NA
ASV507	Veillonellaceae	Veillonella	NA
ASV508	Campylobacteraceae	Campylobacter	NA
ASV509	Lactobacillaceae	Lactobacillus	NA
ASV510	Fusobacteriaceae	Fusobacterium	NA
ASV511	Bifidobacteriaceae	Bifidobacterium	dentium/moukalabense

ASV512	Veillonellaceae	Selenomonas_4	NA
ASV513	NA	NA	NA
ASV514	Fusobacteriaceae	Fusobacterium	NA
ASV515	Spirochaetaceae	Treponema_2	NA
ASV516	Streptococcaceae	Streptococcus	salivarius
ASV517	Prevotellaceae	Alloprevotella	NA
ASV518	Prevotellaceae	Prevotella_7	NA
ASV519	Neisseriaceae	Neisseria	NA
ASV520	Leptotrichiaceae	Leptotrichia	NA
ASV521	Porphyromonadaceae	Porphyromonas	gingivalis
ASV522	Fusobacteriaceae	Fusobacterium	NA
ASV523	Mycoplasmataceae	Mycoplasma	NA
ASV524	Veillonellaceae	Veillonella	NA
ASV525	Fusobacteriaceae	Fusobacterium	NA
ASV526	Desulfovibrionaceae	Desulfovibrio	NA
ASV527	Actinomycetaceae	Actinomyces	massiliensis
ASV528	Prevotellaceae	Prevotella_6	NA
ASV529	Veillonellaceae	Veillonella	NA
ASV530	Veillonellaceae	Veillonella	NA
ASV531	Leptotrichiaceae	Leptotrichia	NA
ASV532	Flavobacteriaceae	Capnocytophaga	NA
ASV533	Prevotellaceae	Alloprevotella	NA
ASV534	Prevotellaceae	Prevotella_7	NA
ASV535	Tannerellaceae	Tannerella	NA
ASV536	Prevotellaceae	Prevotella_6	NA
ASV537	Prevotellaceae	Prevotella_7	buccae

ASV538	Fusobacteriaceae	Fusobacterium	NA
ASV539	Family_XIII	NA	NA
ASV540	Veillonellaceae	Veillonella	NA
ASV541	Prevotellaceae	Prevotella_7	NA
ASV542	Leptotrichiaceae	Leptotrichia	NA
ASV543	Campylobacteraceae	Campylobacter	curvus
ASV544	Prevotellaceae	Alloprevotella	NA
ASV545	Veillonellaceae	Selenomonas_3	NA
ASV546	Actinomycetaceae	Actinomyces	gerencseriae
ASV547	Pasteurellaceae	Haemophilus	NA
ASV548	Fusobacteriaceae	Fusobacterium	NA
ASV549	Lactobacillaceae	Lactobacillus	phage/reuteri/salivarius
ASV550	Prevotellaceae	Prevotella_7	NA
ASV551	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV552	Pasteurellaceae	Haemophilus	NA
ASV553	Streptococcaceae	Streptococcus	NA
ASV554	Pasteurellaceae	Haemophilus	NA
ASV555	Spirochaetaceae	Treponema_2	NA
ASV556	Spirochaetaceae	Treponema_2	NA
ASV557	Prevotellaceae	Prevotella	NA
ASV558	Fusobacteriaceae	Fusobacterium	NA
ASV559	Fusobacteriaceae	Fusobacterium	periodonticum
ASV560	Peptostreptococcaceae	NA	NA
ASV561	Pasteurellaceae	Haemophilus	NA
ASV562	Lachnospiraceae	NA	NA
ASV563	Prevotellaceae	Alloprevotella	NA

ASV564	Campylobacteraceae	Campylobacter	NA
ASV565	Veillonellaceae	Selenomonas_3	NA
ASV566	Flavobacteriaceae	Capnocytophaga	NA
ASV567	Weeksellaceae	Bergeyella	NA
ASV568	Neisseriaceae	Kingella	NA
ASV569	Prevotellaceae	Prevotella_7	pleuritidis
ASV570	Prevotellaceae	Prevotella	NA
ASV571	Spirochaetaceae	Treponema_2	NA
ASV572	Leptotrichiaceae	Leptotrichia	NA
ASV573	Fusobacteriaceae	Fusobacterium	NA
ASV574	Prevotellaceae	Alloprevotella	NA
ASV575	Lactobacillaceae	Lactobacillus	acidophilus/brevis/casei/delbrueckii/fabifermentans/fermentum/helveticus/japonicus/paracasei/paraplantarum/pentosus/plantarum
ASV576	Prevotellaceae	Prevotella_7	NA
ASV577	Neisseriaceae	NA	NA
ASV578	Fusobacteriaceae	Fusobacterium	NA
ASV579	Neisseriaceae	Neisseria	NA
ASV580	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV581	Prevotellaceae	Prevotella_7	NA
ASV582	Prevotellaceae	Prevotella_7	NA
ASV583	Leptotrichiaceae	Leptotrichia	NA
ASV584	Leptotrichiaceae	Leptotrichia	NA
ASV585	Prevotellaceae	Alloprevotella	NA
ASV586	Flavobacteriaceae	Capnocytophaga	ochracea
ASV587	Carnobacteriaceae	Granulicatella	NA
ASV588	Spirochaetaceae	Treponema_2	NA
ASV589	Porphyromonadaceae	Porphyromonas	NA

ASV590	Leptotrichiaceae	Leptotrichia	NA
ASV591	Prevotellaceae	Prevotella_6	NA
ASV592	Lachnospiraceae	NA	NA
ASV593	Prevotellaceae	Prevotella	NA
ASV594	Spirochaetaceae	Treponema_2	NA
ASV595	Leptotrichiaceae	Streptobacillus	NA
ASV596	Flavobacteriaceae	Capnocytophaga	NA
ASV597	Neisseriaceae	Kingella	NA
ASV598	Prevotellaceae	Alloprevotella	NA
ASV599	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV600	Fusobacteriaceae	Fusobacterium	NA
ASV601	Pasteurellaceae	Haemophilus	NA
ASV602	Prevotellaceae	Prevotella_7	NA
ASV603	Lachnospiraceae	Johnsonella	NA
ASV604	Leptotrichiaceae	Leptotrichia	NA
ASV605	Acholeplasmataceae	Acholeplasma	NA
ASV606	Actinomycetaceae	Actinomyces	dentalis/orihominis
ASV607	Veillonellaceae	Anaeroglobus	geminatus
ASV608	Lachnospiraceae	Johnsonella	NA
ASV609	Leptotrichiaceae	Leptotrichia	NA
ASV610	Veillonellaceae	Veillonella	NA
ASV611	Streptococcaceae	Streptococcus	anginosus
ASV612	Veillonellaceae	Dialister	NA
ASV613	Veillonellaceae	Anaeroglobus	NA
ASV614	Prevotellaceae	Alloprevotella	NA
ASV615	Veillonellaceae	Veillonella	NA

ASV616	Cardiobacteriaceae	Cardiobacterium	NA
ASV617	Porphyromonadaceae	Porphyromonas	NA
ASV618	Fusobacteriaceae	Fusobacterium	nucleatum
ASV619	Leptotrichiaceae	Streptobacillus	NA
ASV620	Streptococcaceae	Streptococcus	NA
ASV621	Porphyromonadaceae	Porphyromonas	gingivalis/gulae
ASV622	Veillonellaceae	Veillonella	NA
ASV623	Lachnospiraceae	Johnsonella	NA
ASV624	Weeksellaceae	Bergeyella	NA
ASV625	Pasteurellaceae	Haemophilus	NA
ASV626	Porphyromonadaceae	Porphyromonas	NA
ASV627	Prevotellaceae	Prevotella_7	NA
ASV628	Pasteurellaceae	Actinobacillus	NA
ASV629	Leptotrichiaceae	Leptotrichia	NA
ASV630	Enterobacteriaceae	Klebsiella	NA
ASV631	Spirochaetaceae	Treponema_2	NA
ASV632	Neisseriaceae	NA	NA
ASV633	Fusobacteriaceae	Fusobacterium	NA
ASV634	Veillonellaceae	Veillonella	NA
ASV635	Veillonellaceae	Veillonella	NA
ASV636	Lachnospiraceae	Johnsonella	ignava
ASV637	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV638	Pasteurellaceae	Haemophilus	NA
ASV639	Veillonellaceae	Veillonella	NA
ASV640	Fusobacteriaceae	Fusobacterium	NA
ASV641	Campylobacteraceae	Campylobacter	NA

ASV642	Prevotellaceae	Prevotella_7	NA
ASV643	Pasteurellaceae	Haemophilus	NA
ASV644	Actinomycetaceae	Actinomyces	georgiae/hongkongensis
ASV645	Lachnospiraceae	Johnsonella	NA
ASV646	Prevotellaceae	Alloprevotella	NA
ASV647	Micrococcaceae	Rothia	NA
ASV648	Prevotellaceae	Prevotella_6	NA
ASV649	Prevotellaceae	Alloprevotella	NA
ASV650	Veillonellaceae	Veillonella	NA
ASV651	Prevotellaceae	Alloprevotella	NA
ASV652	Neisseriaceae	Neisseria	NA
ASV653	Prevotellaceae	Prevotella_7	NA
ASV654	Spirochaetaceae	Treponema_2	pectinovorum
ASV655	Pasteurellaceae	Haemophilus	NA
ASV656	Corynebacteriaceae	Corynebacterium	argentoratense/sputi
ASV657	Burkholderiaceae	Comamonas	NA
ASV658	Prevotellaceae	Prevotella_7	NA
ASV659	Weeksellaceae	Bergeyella	NA
ASV660	Weeksellaceae	Bergeyella	NA
ASV661	Prevotellaceae	Prevotella_6	NA
ASV662	Veillonellaceae	Veillonella	NA
ASV663	Prevotellaceae	Prevotella_7	NA
ASV664	Prevotellaceae	Prevotella	NA
ASV665	Lachnospiraceae	NA	NA
ASV666	Prevotellaceae	Prevotella	NA
ASV667	Prevotellaceae	Alloprevotella	NA

ASV668	Veillonellaceae	Veillonella	NA
ASV669	Mycoplasmataceae	Mycoplasma	NA
ASV670	Prevotellaceae	Prevotella	NA
ASV671	Prevotellaceae	Prevotella_7	multisaccharivorax
ASV672	Porphyromonadaceae	Porphyromonas	NA
ASV673	Fusobacteriaceae	Fusobacterium	NA
ASV674	Streptococcaceae	Streptococcus	NA
ASV675	Prevotellaceae	Alloprevotella	NA
ASV676	Weeksellaceae	Bergeyella	NA
ASV677	Streptococcaceae	Streptococcus	NA
ASV678	Prevotellaceae	Prevotella_7	NA
ASV679	Prevotellaceae	Alloprevotella	NA
ASV680	Neisseriaceae	Neisseria	NA
ASV681	Fusobacteriaceae	Fusobacterium	NA
ASV682	Spirochaetaceae	Treponema_2	parvum
ASV683	Neisseriaceae	Kingella	NA
ASV684	Prevotellaceae	Alloprevotella	NA
ASV685	Family_XIII	NA	NA
ASV686	Streptococcaceae	Streptococcus	dentirousetti/downei/sobrinus
ASV687	Actinomycetaceae	Actinomyces	israelii
ASV688	Tannerellaceae	Tannerella	NA
ASV689	Veillonellaceae	Veillonella	NA
ASV690	Veillonellaceae	Veillonella	NA
ASV691	Prevotellaceae	Prevotella_7	buccae/denticola
ASV692	Streptococcaceae	Streptococcus	NA
ASV693	Leptotrichiaceae	Leptotrichia	NA

ASV694	Lachnospiraceae	NA	NA
ASV695	Neisseriaceae	Kingella	NA
ASV696	Campylobacteraceae	Campylobacter	NA
ASV697	Lachnospiraceae	Butyrivibrio_2	NA
ASV698	Lentimicrobiaceae	NA	NA
ASV699	Family_XIII	Family_XIII_UCG-001	NA
ASV700	Neisseriaceae	Kingella	NA
ASV701	Pasteurellaceae	Haemophilus	NA
ASV702	Prevotellaceae	Prevotella_7	NA
ASV703	Pasteurellaceae	Haemophilus	NA
ASV704	Leptotrichiaceae	Leptotrichia	NA
ASV705	Prevotellaceae	Prevotella_7	NA
ASV706	Moraxellaceae	Moraxella	NA
ASV707	Fusobacteriaceae	Fusobacterium	NA
ASV708	NA	NA	NA
ASV709	Fusobacteriaceae	Fusobacterium	NA
ASV710	Prevotellaceae	Alloprevotella	NA
ASV711	Prevotellaceae	Alloprevotella	NA
ASV712	Veillonellaceae	Veillonella	NA
ASV713	NA	NA	NA
ASV714	Bacteroidaceae	Bacteroides	NA
ASV715	Pasteurellaceae	Haemophilus	NA
ASV716	Actinomycetaceae	Mobiluncus	NA
ASV717	Campylobacteraceae	Campylobacter	NA
ASV718	Prevotellaceae	Alloprevotella	NA
ASV719	Fusobacteriaceae	Fusobacterium	NA

ASV720	Neisseriaceae	Kingella	NA
ASV721	Weeksellaceae	Bergeyella	NA
ASV722	NA	NA	NA
ASV723	Micrococcaceae	Rothia	NA
ASV724	Weeksellaceae	Bergeyella	NA
ASV725	Prevotellaceae	Alloprevotella	NA
ASV726	Fusobacteriaceae	Fusobacterium	NA
ASV727	Neisseriaceae	Neisseria	NA
ASV728	Fusobacteriaceae	Fusobacterium	NA
ASV729	Pasteurellaceae	Aggregatibacter	NA
ASV730	Leptotrichiaceae	Leptotrichia	NA
ASV731	Carnobacteriaceae	Granulicatella	NA
ASV732	Carnobacteriaceae	Granulicatella	NA
ASV733	Prevotellaceae	Prevotella_7	NA
ASV734	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV735	Prevotellaceae	Prevotella	NA
ASV736	Prevotellaceae	Prevotella_2	NA
ASV737	NA	NA	NA
ASV738	Fusobacteriaceae	Fusobacterium	NA
ASV739	Prevotellaceae	Alloprevotella	NA
ASV740	Veillonellaceae	Veillonella	NA
ASV741	Leptotrichiaceae	Leptotrichia	trevisanii
ASV742	Fusobacteriaceae	Fusobacterium	NA
ASV743	Leptotrichiaceae	Leptotrichia	NA
ASV744	Neisseriaceae	Neisseria	NA
ASV745	Lachnospiraceae	Oribacterium	NA

ASV746	Weeksellaceae	Bergeyella	NA
ASV747	Lachnospiraceae	Shuttleworthia	satelles
ASV748	Fusobacteriaceae	Fusobacterium	NA
ASV749	Streptococcaceae	Streptococcus	NA
ASV750	Prevotellaceae	Prevotella_7	NA
ASV751	Prevotellaceae	Prevotella_7	NA
ASV752	Prevotellaceae	Prevotella	NA
ASV753	Spirochaetaceae	Treponema_2	NA
ASV754	Veillonellaceae	Veillonella	NA
ASV755	Mitochondria	NA	NA
ASV756	Fusobacteriaceae	Fusobacterium	NA
ASV757	Leptotrichiaceae	Leptotrichia	NA
ASV758	Atopobiaceae	Olsenella	NA
ASV759	Family_XIII	NA	NA
ASV760	Family_XI	Gemella	NA
ASV761	Tannerellaceae	Tannerella	NA
ASV762	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV763	Veillonellaceae	Veillonella	NA
ASV764	Prevotellaceae	Prevotella_7	NA
ASV765	Veillonellaceae	Veillonella	NA
ASV766	Neisseriaceae	Neisseria	NA
ASV767	Lachnospiraceae	Catonella	morbi
ASV768	Neisseriaceae	Neisseria	NA
ASV769	Moraxellaceae	Moraxella	NA
ASV770	Veillonellaceae	NA	NA
ASV771	Lentimicrobiaceae	NA	NA

ASV772	Prevotellaceae	Prevotella_7	NA
ASV773	Prevotellaceae	Prevotella_7	NA
ASV774	Pasteurellaceae	Haemophilus	NA
ASV775	Prevotellaceae	Alloprevotella	NA
ASV776	Erysipelotrichaceae	Solobacterium	NA
ASV777	Porphyromonadaceae	Porphyromonas	NA
ASV778	Prevotellaceae	Prevotella_2	NA
ASV779	Prevotellaceae	Prevotella	NA
ASV780	Prevotellaceae	Alloprevotella	NA
ASV781	Leptotrichiaceae	Leptotrichia	NA
ASV782	Neisseriaceae	Neisseria	NA
ASV783	Prevotellaceae	Prevotella_7	NA
ASV784	Streptococcaceae	Streptococcus	NA
ASV785	Prevotellaceae	Prevotella_7	NA
ASV786	Prevotellaceae	Prevotella_7	NA
ASV787	Fusobacteriaceae	Fusobacterium	NA
ASV788	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV789	Desulfobulbaceae	Desulfobulbus	NA
ASV790	Prevotellaceae	Alloprevotella	NA
ASV791	Porphyromonadaceae	Porphyromonas	uenonis
ASV792	Cardiobacteriaceae	Cardiobacterium	valvarum
ASV793	Pasteurellaceae	Aggregatibacter	NA
ASV794	Cardiobacteriaceae	Cardiobacterium	NA
ASV795	Streptococcaceae	Streptococcus	NA
ASV796	Family_XIII	NA	NA
ASV797	Porphyromonadaceae	Porphyromonas	NA

ASV798	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV799	Family_XI	Gemella	NA
ASV800	Pasteurellaceae	Haemophilus	NA
ASV801	Prevotellaceae	Prevotella_7	enoeca
ASV802	Leptotrichiaceae	Leptotrichia	NA
ASV803	Veillonellaceae	Veillonella	NA
ASV804	Fusobacteriaceae	Fusobacterium	NA
ASV805	Weeksellaceae	Bergeyella	NA
ASV806	Prevotellaceae	Prevotella_7	NA
ASV807	Fusobacteriaceae	Fusobacterium	NA
ASV808	Pasteurellaceae	Haemophilus	NA
ASV809	Cardiobacteriaceae	Cardiobacterium	NA
ASV810	Porphyromonadaceae	Porphyromonas	NA
ASV811	Veillonellaceae	Anaeroglobus	NA
ASV812	Leptotrichiaceae	Leptotrichia	NA
ASV813	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV814	Eggerthellaceae	Cryptobacterium	curtum
ASV815	Spirochaetaceae	Treponema_2	NA
ASV816	Lachnospiraceae	Johnsonella	NA
ASV817	Family_XI	Gemella	NA
ASV818	Micrococcaceae	Rothia	NA
ASV819	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV820	Spirochaetaceae	Treponema_2	NA
ASV821	Neisseriaceae	Neisseria	NA
ASV822	Pasteurellaceae	Actinobacillus	NA
ASV823	Prevotellaceae	Prevotella_7	NA

ASV824	Actinomycetaceae	Actinomyces	NA
ASV825	Prevotellaceae	Prevotella	NA
ASV826	Atopobiaceae	Atopobium	NA
ASV827	NA	NA	NA
ASV828	Family_XIII	Family_XIII_UCG-001	NA
ASV829	Fusobacteriaceae	Fusobacterium	NA
ASV830	Fusobacteriaceae	Fusobacterium	NA
ASV831	Fusobacteriaceae	Fusobacterium	NA
ASV832	Lachnospiraceae	NA	NA
ASV833	Prevotellaceae	Prevotella_6	NA
ASV834	Veillonellaceae	Selenomonas_3	NA
ASV835	Campylobacteraceae	Campylobacter	NA
ASV836	Prevotellaceae	Prevotella	NA
ASV837	Fusobacteriaceae	Fusobacterium	NA
ASV838	Lactobacillaceae	Lactobacillus	antri/fermentum/frumenti/gasseri/oris/panis/reuteri/vaginalis
ASV839	Leptotrichiaceae	Leptotrichia	NA
ASV840	Lachnospiraceae	NA	NA
ASV841	Neisseriaceae	Alysiella	NA
ASV842	Micrococcaceae	Rothia	NA
ASV843	Pasteurellaceae	Haemophilus	NA
ASV844	Fusobacteriaceae	Fusobacterium	NA
ASV845	Spirochaetaceae	Treponema_2	NA
ASV846	Prevotellaceae	Prevotella	NA
ASV847	Veillonellaceae	Veillonella	NA
ASV848	Leptotrichiaceae	Leptotrichia	NA
ASV849	Leptotrichiaceae	Leptotrichia	NA

ASV850	Fusobacteriaceae	Fusobacterium	NA
ASV851	Pasteurellaceae	NA	NA
ASV852	Prevotellaceae	Prevotella_6	NA
ASV853	Leptotrichiaceae	Leptotrichia	NA
ASV854	Neisseriaceae	NA	NA
ASV855	Veillonellaceae	Veillonella	NA
ASV856	Leptotrichiaceae	Leptotrichia	NA
ASV857	Lachnospiraceae	NA	NA
ASV858	Flavobacteriaceae	Capnocytophaga	NA
ASV859	Micrococcaceae	Rothia	NA
ASV860	Fusobacteriaceae	Fusobacterium	NA
ASV861	NA	NA	NA
ASV862	Prevotellaceae	Alloprevotella	NA
ASV863	Prevotellaceae	Alloprevotella	NA
ASV864	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV865	Prevotellaceae	Alloprevotella	NA
ASV866	Pasteurellaceae	NA	NA
ASV867	Spirochaetaceae	Treponema_2	NA
ASV868	Micrococcaceae	Rothia	NA
ASV869	Prevotellaceae	Prevotella_7	NA
ASV870	Prevotellaceae	NA	NA
ASV871	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV872	Prevotellaceae	Alloprevotella	NA
ASV873	Leptotrichiaceae	Leptotrichia	NA
ASV874	Prevotellaceae	Alloprevotella	NA
ASV875	Veillonellaceae	Veillonella	NA

ASV876	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV877	Acholeplasmataceae	Acholeplasma	NA
ASV878	Prevotellaceae	Prevotella_7	NA
ASV879	Prevotellaceae	Prevotella_6	NA
ASV880	Carnobacteriaceae	Granulicatella	NA
ASV881	Leptotrichiaceae	Leptotrichia	NA
ASV882	Mycoplasmataceae	Mycoplasma	NA
ASV883	Lactobacillaceae	Lactobacillus	brevis/casei/pantheris/rhamnosus
ASV884	Veillonellaceae	Veillonella	NA
ASV885	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV886	Leptotrichiaceae	Leptotrichia	NA
ASV887	Lactobacillaceae	Lactobacillus	oris
ASV888	Neisseriaceae	Neisseria	NA
ASV889	Pasteurellaceae	Haemophilus	NA
ASV890	Prevotellaceae	Prevotella_6	NA
ASV891	Pasteurellaceae	Haemophilus	NA
ASV892	Veillonellaceae	Megasphaera	NA
ASV893	Family_XI	Parvimonas	micra
ASV894	Veillonellaceae	Veillonella	NA
ASV895	Veillonellaceae	Dialister	NA
ASV896	Leptotrichiaceae	Leptotrichia	NA
ASV897	Lachnospiraceae	Lachnoanaerobaculum	umeaense
ASV898	Fusobacteriaceae	Fusobacterium	NA
ASV899	Actinomycetaceae	Actinomyces	NA
ASV900	Campylobacteraceae	Campylobacter	NA
ASV901	Prevotellaceae	Prevotella	NA

ASV902	Prevotellaceae	Alloprevotella	NA
ASV903	Veillonellaceae	Selenomonas_3	NA
ASV904	Bacteroidaceae	Bacteroides	heparinolyticus
ASV905	Fusobacteriaceae	Fusobacterium	NA
ASV906	Fusobacteriaceae	Fusobacterium	NA
ASV907	Streptococcaceae	Streptococcus	NA
ASV908	Actinomycetaceae	NA	NA
ASV909	Prevotellaceae	Prevotella_7	NA
ASV910	Desulfovibrionaceae	Desulfovibrio	NA
ASV911	Fusobacteriaceae	Fusobacterium	NA
ASV912	Leptotrichiaceae	Leptotrichia	NA
ASV913	Prevotellaceae	Prevotella_7	NA
ASV914	Fusobacteriaceae	Fusobacterium	NA
ASV915	Prevotellaceae	Prevotella_7	NA
ASV916	Neisseriaceae	Neisseria	NA
ASV917	Veillonellaceae	Veillonella	NA
ASV918	Prevotellaceae	Alloprevotella	NA
ASV919	Prevotellaceae	Prevotella_7	NA
ASV920	Streptococcaceae	Streptococcus	NA
ASV921	Spirochaetaceae	Treponema_2	NA
ASV922	Prevotellaceae	Prevotella_6	NA
ASV923	Lachnospiraceae	Stomatobaculum	NA
ASV924	Pasteurellaceae	Aggregatibacter	NA
ASV925	Streptococcaceae	Streptococcus	NA
ASV926	Actinomycetaceae	Actinomyces	meyeri
ASV927	Veillonellaceae	Veillonella	NA

ASV928	Leptotrichiaceae	Leptotrichia	NA
ASV929	Spirochaetaceae	Treponema_2	NA
ASV930	Pasteurellaceae	Haemophilus	NA
ASV931	Porphyromonadaceae	Porphyromonas	NA
ASV932	Veillonellaceae	Veillonella	NA
ASV933	Prevotellaceae	Prevotella_6	NA
ASV934	Fusobacteriaceae	Fusobacterium	NA
ASV935	Fusobacteriaceae	Fusobacterium	NA
ASV936	Veillonellaceae	Veillonella	NA
ASV937	Carnobacteriaceae	Granulicatella	NA
ASV938	Prevotellaceae	Alloprevotella	NA
ASV939	Leptotrichiaceae	Leptotrichia	NA
ASV940	Lachnospiraceae	Oribacterium	NA
ASV941	Veillonellaceae	Veillonella	NA
ASV942	Neisseriaceae	Neisseria	NA
ASV943	Prevotellaceae	Prevotella_7	NA
ASV944	Lentimicrobiaceae	NA	NA
ASV945	Campylobacteraceae	Campylobacter	NA
ASV946	Leptotrichiaceae	Leptotrichia	NA
ASV947	Prevotellaceae	Prevotella_7	NA
ASV948	Fusobacteriaceae	Fusobacterium	NA
ASV949	Streptococcaceae	Streptococcus	NA
ASV950	Pasteurellaceae	Aggregatibacter	NA
ASV951	Actinomycetaceae	Actinomyces	cardiffensis
ASV952	Leptotrichiaceae	Leptotrichia	NA
ASV953	Fusobacteriaceae	Fusobacterium	NA

ASV954	Pasteurellaceae	Rodentibacter	NA
ASV955	Cardiobacteriaceae	Cardiobacterium	NA
ASV956	Prevotellaceae	Prevotella_7	NA
ASV957	Prevotellaceae	Alloprevotella	NA
ASV958	Veillonellaceae	Selenomonas	NA
ASV959	Prevotellaceae	Prevotella_7	NA
ASV960	Lactobacillaceae	Lactobacillus	NA
ASV961	Leptotrichiaceae	Leptotrichia	NA
ASV962	Leptotrichiaceae	Leptotrichia	NA
ASV963	Family_XI	Gemella	NA
ASV964	Spirochaetaceae	Treponema_2	NA
ASV965	Desulfovibrionaceae	Bilophila	wadsworthia
ASV966	Family_XI	Gemella	NA
ASV967	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV968	Leptotrichiaceae	Leptotrichia	NA
ASV969	Pasteurellaceae	Haemophilus	NA
ASV970	Spirochaetaceae	Treponema_2	NA
ASV971	Neisseriaceae	Kingella	kingae
ASV972	Neisseriaceae	Neisseria	NA
ASV973	Burkholderiaceae	Bordetella	holmesii/pertussis
ASV974	Veillonellaceae	Veillonella	NA
ASV975	Pasteurellaceae	Haemophilus	NA
ASV976	Prevotellaceae	Prevotella_7	NA
ASV977	Pasteurellaceae	Haemophilus	NA
ASV978	Lachnospiraceae	Johnsonella	NA
ASV979	Veillonellaceae	Veillonella	NA

ASV980	Pasteurellaceae	Haemophilus	NA
ASV981	Lentimicrobiaceae	NA	NA
ASV982	Porphyromonadaceae	Porphyromonas	NA
ASV983	Campylobacteraceae	Campylobacter	NA
ASV984	Prevotellaceae	Prevotella	NA
ASV985	NA	NA	NA
ASV986	Veillonellaceae	Veillonella	NA
ASV987	Leptotrichiaceae	Leptotrichia	NA
ASV988	Veillonellaceae	Veillonella	NA
ASV989	Prevotellaceae	Prevotella_7	NA
ASV990	Prevotellaceae	Alloprevotella	NA
ASV991	Carnobacteriaceae	Granulicatella	NA
ASV992	Leptotrichiaceae	Leptotrichia	NA
ASV993	Campylobacteraceae	Campylobacter	NA
ASV994	Carnobacteriaceae	NA	NA
ASV995	Desulfomicrobiaceae	Desulfomicrobium	orale
ASV996	Peptostreptococcaceae	Peptoanaerobacter	NA
ASV997	Lachnospiraceae	NA	NA
ASV998	Leptotrichiaceae	Leptotrichia	NA
ASV999	Prevotellaceae	Prevotella_7	NA
ASV1000	Pasteurellaceae	Aggregatibacter	NA
ASV1001	Leptotrichiaceae	Leptotrichia	NA
ASV1002	Prevotellaceae	Prevotella	NA
ASV1003	Fusobacteriaceae	Fusobacterium	NA
ASV1004	NA	NA	NA
ASV1005	Veillonellaceae	Allisonella	NA

ASV1006	Veillonellaceae	Veillonella	NA
ASV1007	Porphyromonadaceae	Porphyromonas	NA
ASV1008	Pasteurellaceae	Aggregatibacter	NA
ASV1009	NA	NA	NA
ASV1010	Actinomycetaceae	Actinomyces	NA
ASV1011	Leptotrichiaceae	Leptotrichia	NA
ASV1012	Fusobacteriaceae	Fusobacterium	nucleatum
ASV1013	Spirochaetaceae	Treponema_2	NA
ASV1014	Lachnospiraceae	Johnsonella	NA
ASV1015	Neisseriaceae	Neisseria	NA
ASV1016	Neisseriaceae	Neisseria	NA
ASV1017	Weeksellaceae	Bergeyella	NA
ASV1018	Leptotrichiaceae	Leptotrichia	NA
ASV1019	Flavobacteriaceae	Capnocytophaga	NA
ASV1020	Neisseriaceae	Kingella	NA
ASV1021	Pasteurellaceae	Aggregatibacter	NA
ASV1022	Fusobacteriaceae	Fusobacterium	NA
ASV1023	Fusobacteriaceae	Fusobacterium	NA
ASV1024	Porphyromonadaceae	Porphyromonas	NA
ASV1025	Carnobacteriaceae	Granulicatella	NA
ASV1026	Lentimicrobiaceae	NA	NA
ASV1027	Prevotellaceae	Prevotella_7	NA
ASV1028	Leptotrichiaceae	Leptotrichia	NA
ASV1029	Pasteurellaceae	Mannheimia	NA
ASV1030	Neisseriaceae	Alysiella	NA
ASV1031	Veillonellaceae	Veillonella	NA

ASV1032	Leptotrichiaceae	Leptotrichia	NA
ASV1033	Flavobacteriaceae	Capnocytophaga	NA
ASV1034	Acholeplasmataceae	Acholeplasma	NA
ASV1035	Lachnospiraceae	Stomatobaculum	NA
ASV1036	Neisseriaceae	Neisseria	NA
ASV1037	Prevotellaceae	Prevotella_7	NA
ASV1038	Prevotellaceae	Prevotella_7	NA
ASV1039	Prevotellaceae	Prevotella_6	NA
ASV1040	Fusobacteriaceae	Fusobacterium	NA
ASV1041	Prevotellaceae	Prevotella_7	NA
ASV1042	Lachnospiraceae	Johnsonella	NA
ASV1043	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1044	Prevotellaceae	Prevotella_2	NA
ASV1045	Porphyromonadaceae	Porphyromonas	NA
ASV1046	Leptotrichiaceae	Leptotrichia	NA
ASV1047	Prevotellaceae	Prevotella_7	NA
ASV1048	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1049	Veillonellaceae	Veillonella	NA
ASV1050	Tannerellaceae	Tannerella	NA
ASV1051	Prevotellaceae	Prevotella_7	NA
ASV1052	Prevotellaceae	Prevotella_7	NA
ASV1053	Prevotellaceae	Alloprevotella	NA
ASV1054	Pasteurellaceae	Haemophilus	influenzae
ASV1055	Campylobacteraceae	Campylobacter	NA
ASV1056	Prevotellaceae	Prevotella	NA
ASV1057	Lachnospiraceae	Lachnoanaerobaculum	NA

ASV1058	Family_XIII	NA	NA
ASV1059	Fusobacteriaceae	Fusobacterium	NA
ASV1060	Pasteurellaceae	Aggregatibacter	NA
ASV1061	Leptotrichiaceae	Leptotrichia	NA
ASV1062	Veillonellaceae	Veillonella	NA
ASV1063	Fusobacteriaceae	Fusobacterium	NA
ASV1064	Pasteurellaceae	Haemophilus	NA
ASV1065	Campylobacteraceae	Campylobacter	NA
ASV1066	Veillonellaceae	Veillonella	NA
ASV1067	Veillonellaceae	Veillonella	NA
ASV1068	Fusobacteriaceae	Fusobacterium	NA
ASV1069	Bacteroidaceae	Bacteroides	NA
ASV1070	Lactobacillaceae	Lactobacillus	acidophilus/casei/crispatus/gallinarum/helveticus/kitasatonis
ASV1071	Neisseriaceae	NA	NA
ASV1072	Pasteurellaceae	Haemophilus	haemolyticus/influenzae
ASV1073	Carnobacteriaceae	Granulicatella	NA
ASV1074	Leptotrichiaceae	Leptotrichia	NA
ASV1075	Porphyromonadaceae	Porphyromonas	NA
ASV1076	Prevotellaceae	Prevotella_7	NA
ASV1077	Prevotellaceae	Prevotella_6	NA
ASV1078	Flavobacteriaceae	Capnocytophaga	NA
ASV1079	Veillonellaceae	Veillonella	NA
ASV1080	Spirochaetaceae	Treponema_2	NA
ASV1081	NA	NA	NA
ASV1082	Streptococcaceae	Streptococcus	NA
ASV1083	Leptotrichiaceae	Leptotrichia	NA

ASV1084	Spirochaetaceae	Treponema_2	NA
ASV1085	Actinomycetaceae	Actinomyces	NA
ASV1086	Leptotrichiaceae	Leptotrichia	NA
ASV1087	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV1088	Veillonellaceae	Dialister	micraerophilus/microaerophilus
ASV1089	Carnobacteriaceae	Granulicatella	NA
ASV1090	Lachnospiraceae	Stomatobaculum	NA
ASV1091	Veillonellaceae	Selenomonas	NA
ASV1092	Porphyromonadaceae	Porphyromonas	NA
ASV1093	Spirochaetaceae	Treponema_2	NA
ASV1094	Veillonellaceae	Veillonella	NA
ASV1095	Veillonellaceae	Selenomonas_3	NA
ASV1096	Atopobiaceae	Atopobium	NA
ASV1097	Leptotrichiaceae	Leptotrichia	NA
ASV1098	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1099	Fusobacteriaceae	Fusobacterium	NA
ASV1100	Prevotellaceae	Prevotella_7	NA
ASV1101	Weeksellaceae	Bergeyella	NA
ASV1102	Neisseriaceae	Neisseria	NA
ASV1103	Fusobacteriaceae	Fusobacterium	NA
ASV1104	Prevotellaceae	Prevotella_7	NA
ASV1105	Porphyromonadaceae	Porphyromonas	NA
ASV1106	Prevotellaceae	Prevotella_7	NA
ASV1107	Prevotellaceae	Prevotella	NA
ASV1108	Weeksellaceae	Bergeyella	NA
ASV1109	Prevotellaceae	Prevotella	NA

ASV1110	Veillonellaceae	Veillonella	NA
ASV1111	Leptotrichiaceae	Leptotrichia	NA
ASV1112	Veillonellaceae	Veillonella	NA
ASV1113	Synergistaceae	Fretibacterium	NA
ASV1114	Veillonellaceae	Veillonella	NA
ASV1115	Leptotrichiaceae	Leptotrichia	NA
ASV1116	Atopobiaceae	Atopobium	NA
ASV1117	Veillonellaceae	Veillonella	NA
ASV1118	Prevotellaceae	Prevotella_2	NA
ASV1119	Pasteurellaceae	Mannheimia	NA
ASV1120	Lachnospiraceae	Johnsonella	NA
ASV1121	Fusobacteriaceae	Fusobacterium	NA
ASV1122	Lentimicrobiaceae	NA	NA
ASV1123	Fusobacteriaceae	Fusobacterium	NA
ASV1124	Prevotellaceae	Prevotella_7	NA
ASV1125	Eggerthellaceae	Slackia	exigua
ASV1126	Prevotellaceae	Alloprevotella	NA
ASV1127	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia	NA
ASV1128	Prevotellaceae	Prevotella_7	NA
ASV1129	Veillonellaceae	Veillonella	NA
ASV1130	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV1131	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1132	Fusobacteriaceae	Fusobacterium	NA
ASV1133	NA	NA	NA
ASV1134	Fusobacteriaceae	Fusobacterium	NA

ASV1135	Helicobacteraceae	Wolinella	NA
ASV1136	Prevotellaceae	Prevotella	NA
ASV1137	Prevotellaceae	Alloprevotella	NA
ASV1138	Neisseriaceae	Conchiformibius	NA
ASV1139	Neisseriaceae	NA	NA
ASV1140	Prevotellaceae	Prevotella	NA
ASV1141	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV1142	Spirochaetaceae	Treponema_2	NA
ASV1143	Prevotellaceae	Prevotella_6	NA
ASV1144	Neisseriaceae	Kingella	NA
ASV1145	p-2534-18B5_gut_group	NA	NA
ASV1146	Prevotellaceae	Prevotella_7	NA
ASV1147	Fusobacteriaceae	Fusobacterium	NA
ASV1148	Spirochaetaceae	Treponema_2	NA
ASV1149	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV1150	Family_XI	Parvimonas	NA
ASV1151	Campylobacteraceae	Campylobacter	NA
ASV1152	Prevotellaceae	Alloprevotella	NA
ASV1153	Fusobacteriaceae	Fusobacterium	NA
ASV1154	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1155	Fusobacteriaceae	Fusobacterium	NA
ASV1156	Veillonellaceae	Veillonella	NA
ASV1157	Pasteurellaceae	Haemophilus	NA
ASV1158	Fusobacteriaceae	Fusobacterium	NA
ASV1159	Eubacteriaceae	Pseudoramibacter	alactolyticus
ASV1160	Fusobacteriaceae	Fusobacterium	NA

ASV1161	Prevotellaceae	Alloprevotella	NA
ASV1162	Lachnospiraceae	NA	NA
ASV1163	Prevotellaceae	Prevotella_7	NA
ASV1164	Leptotrichiaceae	Leptotrichia	NA
ASV1165	Prevotellaceae	Alloprevotella	NA
ASV1166	Leptotrichiaceae	Leptotrichia	NA
ASV1167	Burkholderiaceae	Comamonas	NA
ASV1168	Fusobacteriaceae	Fusobacterium	NA
ASV1169	Spirochaetaceae	Treponema_2	NA
ASV1170	Fusobacteriaceae	Fusobacterium	NA
ASV1171	NA	NA	NA
ASV1172	Paludibacteraceae	F0058	NA
ASV1173	Neisseriaceae	Neisseria	NA
ASV1174	Coxiellaceae	Coxiella	NA
ASV1175	Micrococcaceae	Rothia	NA
ASV1176	NA	NA	NA
ASV1177	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV1178	Fusobacteriaceae	Fusobacterium	NA
ASV1179	Prevotellaceae	Prevotella_7	NA
ASV1180	Prevotellaceae	NA	NA
ASV1181	Fusobacteriaceae	Fusobacterium	NA
ASV1182	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1183	Prevotellaceae	Prevotella	NA
ASV1184	Prevotellaceae	Prevotella_7	NA
ASV1185	Prevotellaceae	Prevotella_7	NA
ASV1186	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA

ASV1187	Spirochaetaceae	Treponema_2	putidum
ASV1188	Atopobiaceae	Olsenella	NA
ASV1189	Prevotellaceae	Alloprevotella	NA
ASV1190	Porphyromonadaceae	Porphyromonas	NA
ASV1191	Burkholderiaceae	NA	NA
ASV1192	Lachnospiraceae	Butyrivibrio_2	NA
ASV1193	Veillonellaceae	Selenomonas	NA
ASV1194	Prevotellaceae	Prevotella_7	NA
ASV1195	Synergistaceae	Fretibacterium	NA
ASV1196	Fusobacteriaceae	Fusobacterium	NA
ASV1197	Family_XIII	NA	NA
ASV1198	Weeksellaceae	Bergeyella	NA
ASV1199	Veillonellaceae	Veillonella	NA
ASV1200	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1201	Spirochaetaceae	Treponema_2	NA
ASV1202	Prevotellaceae	Prevotella_6	NA
ASV1203	Flavobacteriaceae	Capnocytophaga	NA
ASV1204	Campylobacteraceae	Campylobacter	NA
ASV1205	Acholeplasmataceae	Acholeplasma	NA
ASV1206	Campylobacteraceae	Campylobacter	NA
ASV1207	Spirochaetaceae	Treponema_2	NA
ASV1208	Pasteurellaceae	Aggregatibacter	NA
ASV1209	Family_XIII	NA	NA
ASV1210	Lachnospiraceae	Stomatobaculum	NA
ASV1211	Streptococcaceae	Streptococcus	NA
ASV1212	Streptococcaceae	Streptococcus	NA

ASV1213	Veillonellaceae	Selenomonas_3	NA
ASV1214	Micrococcaceae	Rothia	NA
ASV1215	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1216	Streptococcaceae	Streptococcus	NA
ASV1217	Prevotellaceae	Alloprevotella	NA
ASV1218	Flavobacteriaceae	Capnocytophaga	NA
ASV1219	Fusobacteriaceae	Fusobacterium	NA
ASV1220	Veillonellaceae	Dialister	NA
ASV1221	Flavobacteriaceae	Capnocytophaga	NA
ASV1222	Prevotellaceae	Prevotella_7	NA
ASV1223	Spirochaetaceae	Treponema_2	NA
ASV1224	Leptotrichiaceae	Leptotrichia	NA
ASV1225	Prevotellaceae	Alloprevotella	NA
ASV1226	Prevotellaceae	Prevotella	NA
ASV1227	Cardiobacteriaceae	Cardiobacterium	NA
ASV1228	Synergistaceae	Pyramidobacter	piscolens
ASV1229	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1230	Prevotellaceae	Alloprevotella	NA
ASV1231	Spirochaetaceae	Treponema_2	NA
ASV1232	Veillonellaceae	Veillonella	NA
ASV1233	Spirochaetaceae	Treponema_2	NA
ASV1234	Pasteurellaceae	Actinobacillus	NA
ASV1235	Porphyromonadaceae	Porphyromonas	NA
ASV1236	Fusobacteriaceae	Fusobacterium	NA
ASV1237	Lactobacillaceae	Lactobacillus	NA
ASV1238	Prevotellaceae	Prevotella_2	NA

ASV1239	Spirochaetaceae	Treponema_2	NA
ASV1240	Flavobacteriaceae	Capnocytophaga	NA
ASV1241	NA	NA	NA
ASV1242	Prevotellaceae	Prevotella	NA
ASV1243	Streptococcaceae	Streptococcus	NA
ASV1244	Weeksellaceae	Bergeyella	NA
ASV1245	Family_XIII	Mogibacterium	NA
ASV1246	Erysipelotrichaceae	Eggerthia	catenaformis
ASV1247	Streptococcaceae	Streptococcus	NA
ASV1248	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1249	Flavobacteriaceae	Capnocytophaga	NA
ASV1250	Family_XI	Gemella	NA
ASV1251	Streptococcaceae	Streptococcus	NA
ASV1252	Veillonellaceae	Veillonella	NA
ASV1253	Weeksellaceae	Bergeyella	NA
ASV1254	Streptococcaceae	Streptococcus	NA
ASV1255	Lachnospiraceae	Stomatobaculum	NA
ASV1256	Prevotellaceae	Prevotella_7	NA
ASV1257	Fusobacteriaceae	Fusobacterium	NA
ASV1258	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1259	Campylobacteraceae	Campylobacter	NA
ASV1260	Family_XIII	Mogibacterium	NA
ASV1261	Lachnospiraceae	Stomatobaculum	NA
ASV1262	Prevotellaceae	Alloprevotella	NA
ASV1263	Streptococcaceae	Streptococcus	NA
ASV1264	Lactobacillaceae	Lactobacillus	acidophilus/amylovorus/ultunensis

ASV1265	NA	NA	NA
ASV1266	Weeksellaceae	Bergeyella	cardium
ASV1267	Spirochaetaceae	Treponema_2	NA
ASV1268	Prevotellaceae	Prevotella	NA
ASV1269	Leptotrichiaceae	Streptobacillus	NA
ASV1270	Neisseriaceae	Neisseria	NA
ASV1271	Prevotellaceae	Alloprevotella	NA
ASV1272	Prevotellaceae	Prevotella_7	NA
ASV1273	Neisseriaceae	Alysiella	NA
ASV1274	Carnobacteriaceae	Granulicatella	NA
ASV1275	Streptococcaceae	Streptococcus	NA
ASV1276	Family_XI	Parvimonas	NA
ASV1277	Prevotellaceae	Prevotella	NA
ASV1278	Lactobacillaceae	Lactobacillus	paraplantarum/plantarum
ASV1279	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1280	Prevotellaceae	Alloprevotella	NA
ASV1281	Lactobacillaceae	Lactobacillus	amylovorus
ASV1282	Lactobacillaceae	Lactobacillus	brevis/buchneri/helveticus/kefiri/otakiensis/parabuchneri/pontis/sunkii
ASV1283	Fusobacteriaceae	Fusobacterium	NA
ASV1284	Neisseriaceae	Neisseria	NA
ASV1285	Leptotrichiaceae	Leptotrichia	NA
ASV1286	Burkholderiaceae	Pelomonas	aquatica/puraquae/saccharophila
ASV1287	Prevotellaceae	Prevotella_7	NA
ASV1288	Spirochaetaceae	Treponema_2	NA
ASV1289	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1290	Campylobacteraceae	Campylobacter	NA

ASV1291	Prevotellaceae	Alloprevotella	NA
ASV1292	Neisseriaceae	Alysiella	NA
ASV1293	Prevotellaceae	Prevotella_7	NA
ASV1294	NA	NA	NA
ASV1295	Actinomycetaceae	Actinomyces	NA
ASV1296	Acholeplasmataceae	Acholeplasma	NA
ASV1297	Prevotellaceae	Alloprevotella	NA
ASV1298	Prevotellaceae	Alloprevotella	NA
ASV1299	Neisseriaceae	NA	NA
ASV1300	Fusobacteriaceae	Fusobacterium	NA
ASV1301	Carnobacteriaceae	Granulicatella	NA
ASV1302	Family_XI	Gemella	NA
ASV1303	Prevotellaceae	Prevotella	NA
ASV1304	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1305	Prevotellaceae	Prevotella	NA
ASV1306	Pasteurellaceae	NA	NA
ASV1307	Prevotellaceae	Prevotella_7	NA
ASV1308	Pasteurellaceae	Haemophilus	NA
ASV1309	Streptococcaceae	Streptococcus	NA
ASV1310	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1311	Fusobacteriaceae	Fusobacterium	NA
ASV1312	Atopobiaceae	Olsenella	uli
ASV1313	Family_XIII	Mogibacterium	NA
ASV1314	Veillonellaceae	Veillonella	NA
ASV1315	Lactobacillaceae	Lactobacillus	NA
ASV1316	Xanthobacteraceae	NA	NA

ASV1317	Prevotellaceae	Prevotella	NA
ASV1318	Leptotrichiaceae	Leptotrichia	NA
ASV1319	Leptotrichiaceae	Leptotrichia	NA
ASV1320	Streptococcaceae	Streptococcus	NA
ASV1321	Flavobacteriaceae	Capnocytophaga	NA
ASV1322	Leptotrichiaceae	Leptotrichia	NA
ASV1323	Neisseriaceae	Kingella	NA
ASV1324	Weeksellaceae	Bergeyella	NA
ASV1325	Neisseriaceae	Kingella	NA
ASV1326	Veillonellaceae	Selenomonas_3	NA
ASV1327	Lentimicrobiaceae	NA	NA
ASV1328	Neisseriaceae	Neisseria	NA
ASV1329	Prevotellaceae	Prevotella	NA
ASV1330	Neisseriaceae	Neisseria	NA
ASV1331	Veillonellaceae	Selenomonas_3	NA
ASV1332	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1333	Lachnospiraceae	Howardella	NA
ASV1334	Neisseriaceae	Neisseria	NA
ASV1335	Fusobacteriaceae	Fusobacterium	NA
ASV1336	Micrococcaceae	Rothia	NA
ASV1337	Lachnospiraceae	Shuttleworthia	satelles
ASV1338	Lachnospiraceae	Stomatobaculum	NA
ASV1339	Carnobacteriaceae	Granulicatella	NA
ASV1340	Prevotellaceae	Prevotella_7	NA
ASV1341	Fusobacteriaceae	Fusobacterium	NA
ASV1342	Neisseriaceae	NA	NA

ASV1343	Spirochaetaceae	Treponema_2	NA
ASV1344	Prevotellaceae	Prevotella_7	NA
ASV1345	Prevotellaceae	Prevotella	NA
ASV1346	Prevotellaceae	Prevotella_6	NA
ASV1347	Fusobacteriaceae	Fusobacterium	NA
ASV1348	Prevotellaceae	Alloprevotella	NA
ASV1349	Prevotellaceae	Prevotella	NA
ASV1350	Streptococcaceae	Streptococcus	NA
ASV1351	Fusobacteriaceae	Fusobacterium	NA
ASV1352	Veillonellaceae	Veillonella	NA
ASV1353	Prevotellaceae	NA	NA
ASV1354	Lachnospiraceae	Oribacterium	NA
ASV1355	Lachnospiraceae	Stomatobaculum	NA
ASV1356	Paludibacteraceae	F0058	NA
ASV1357	Prevotellaceae	Prevotella	NA
ASV1358	Streptococcaceae	Streptococcus	NA
ASV1359	Carnobacteriaceae	Granulicatella	NA
ASV1360	Prevotellaceae	Alloprevotella	NA
ASV1361	Moraxellaceae	Acinetobacter	baumannii/bouvetii/haemolyticus/johnsonii/junii/lwoffii/oleivorans/oryzae/schindleri
ASV1362	Prevotellaceae	Prevotella_7	NA
ASV1363	Spirochaetaceae	Treponema_2	NA
ASV1364	Lachnospiraceae	Stomatobaculum	NA
ASV1365	Fusobacteriaceae	Fusobacterium	NA
ASV1366	Fusobacteriaceae	Fusobacterium	NA
ASV1367	Muribaculaceae	NA	NA
ASV1368	Lachnospiraceae	NA	NA

ASV1369	Leptotrichiaceae	Leptotrichia	NA
ASV1370	Veillonellaceae	Veillonella	NA
ASV1371	Streptococcaceae	Streptococcus	NA
ASV1372	Prevotellaceae	Prevotella	NA
ASV1373	Lachnospiraceae	Catonella	NA
ASV1374	Spirochaetaceae	Treponema_2	NA
ASV1375	Leptotrichiaceae	Leptotrichia	NA
ASV1376	Leptotrichiaceae	Leptotrichia	NA
ASV1377	Carnobacteriaceae	Granulicatella	NA
ASV1378	Bifidobacteriaceae	Alloscardovia	NA
ASV1379	NA	NA	NA
ASV1380	Spirochaetaceae	Treponema_2	NA
ASV1381	Lachnospiraceae	Oribacterium	NA
ASV1382	Fusobacteriaceae	Fusobacterium	NA
ASV1383	Cardiobacteriaceae	NA	NA
ASV1384	Leptotrichiaceae	Leptotrichia	NA
ASV1385	Actinomycetaceae	Actinomyces	NA
ASV1386	Family_XIII	NA	NA
ASV1387	Veillonellaceae	Megasphaera	NA
ASV1388	Veillonellaceae	Veillonella	NA
ASV1389	Actinomycetaceae	Actinomyces	NA
ASV1390	Prevotellaceae	Prevotella_6	NA
ASV1391	Spirochaetaceae	Treponema_2	NA
ASV1392	Lachnospiraceae	Oribacterium	NA
ASV1393	Weeksellaceae	Bergeyella	NA
ASV1394	Corynebacteriaceae	Corynebacterium_1	aurimucosum/pseudogenitalium/tuberculostearicum

ASV1395	Prevotellaceae	Alloprevotella	NA
ASV1396	Campylobacteraceae	Campylobacter	NA
ASV1397	Fusobacteriaceae	Fusobacterium	NA
ASV1398	Campylobacteraceae	Campylobacter	NA
ASV1399	Neisseriaceae	Kingella	NA
ASV1400	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1401	Fusobacteriaceae	Fusobacterium	NA
ASV1402	Lachnospiraceae	Stomatobaculum	NA
ASV1403	Cardiobacteriaceae	Cardiobacterium	NA
ASV1404	Veillonellaceae	Veillonella	NA
ASV1405	Lachnospiraceae	Stomatobaculum	NA
ASV1406	Carnobacteriaceae	Granulicatella	NA
ASV1407	Leptotrichiaceae	Leptotrichia	NA
ASV1408	Family_XI	Gemella	NA
ASV1409	Moraxellaceae	Acinetobacter	baumannii/beijerinckii/bereziniae/calcoaceticus/courvalinii/gerneri/guillouiae/gyllenbergii/indicus/johnsonii/junii/lwoffii/modestus/parvus/plantaru
ASV1410	Flavobacteriaceae	Capnocytophaga	NA
ASV1411	Actinomycetaceae	Actinomyces	NA
ASV1412	Carnobacteriaceae	Granulicatella	NA
ASV1413	Veillonellaceae	Veillonella	NA
ASV1414	Actinomycetaceae	Actinomyces	NA
ASV1415	Lachnospiraceae	Howardella	NA
ASV1416	NA	NA	NA
ASV1417	Flavobacteriaceae	Capnocytophaga	NA
ASV1418	Pasteurellaceae	Haemophilus	NA
ASV1419	Leptotrichiaceae	Leptotrichia	NA
ASV1420	Leptotrichiaceae	Leptotrichia	NA

ASV1421	NA	NA	NA
ASV1422	Flavobacteriaceae	Capnocytophaga	NA
ASV1423	Family_XI	Gemella	NA
ASV1424	Veillonellaceae	Veillonella	NA
ASV1425	Porphyromonadaceae	Porphyromonas	NA
ASV1426	Fusobacteriaceae	Fusobacterium	NA
ASV1427	NA	NA	NA
ASV1428	Prevotellaceae	Alloprevotella	NA
ASV1429	Veillonellaceae	Veillonella	NA
ASV1430	Prevotellaceae	Alloprevotella	NA
ASV1431	Prevotellaceae	Prevotella_7	NA
ASV1432	Leptotrichiaceae	Leptotrichia	NA
ASV1433	Porphyromonadaceae	Porphyromonas	NA
ASV1434	Prevotellaceae	Prevotella_7	NA
ASV1435	Prevotellaceae	Prevotella_7	NA
ASV1436	Prevotellaceae	NA	NA
ASV1437	Flavobacteriaceae	Capnocytophaga	NA
ASV1438	Neisseriaceae	Neisseria	NA
ASV1439	Streptococcaceae	Streptococcus	NA
ASV1440	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1441	Pasteurellaceae	Haemophilus	NA
ASV1442	Spirochaetaceae	Treponema_2	NA
ASV1443	Neisseriaceae	NA	NA
ASV1444	Enterobacteriaceae	Raoultella	electrica/ornithinolytica/planticola
ASV1445	Leptotrichiaceae	Streptobacillus	hongkongensis
ASV1446	Erysipelotrichaceae	Erysipelotrichaceae_UCG-004	NA

ASV1447	Lachnospiraceae	NA	NA
ASV1448	Fusobacteriaceae	Fusobacterium	NA
ASV1449	Prevotellaceae	Prevotella_7	NA
ASV1450	Peptostreptococcaceae	Peptoanaerobacter	NA
ASV1451	Fusobacteriaceae	Fusobacterium	NA
ASV1452	Prevotellaceae	Prevotella_6	NA
ASV1453	Spirochaetaceae	Treponema_2	NA
ASV1454	Porphyromonadaceae	Porphyromonas	NA
ASV1455	Fusobacteriaceae	Fusobacterium	NA
ASV1456	Flavobacteriaceae	Capnocytophaga	NA
ASV1457	Spirochaetaceae	Treponema_2	NA
ASV1458	Peptostreptococcaceae	NA	NA
ASV1459	Weeksellaceae	Bergeyella	NA
ASV1460	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1461	Prevotellaceae	Alloprevotella	NA
ASV1462	Lachnospiraceae	Oribacterium	NA
ASV1463	Leptotrichiaceae	Leptotrichia	NA
ASV1464	Prevotellaceae	Prevotella	NA
ASV1465	Prevotellaceae	Prevotella_7	NA
ASV1466	NA	NA	NA
ASV1467	Prevotellaceae	Prevotella_6	NA
ASV1468	Neisseriaceae	Kingella	NA
ASV1469	NA	NA	NA
ASV1470	Streptococcaceae	Streptococcus	NA
ASV1471	Fusobacteriaceae	Fusobacterium	NA
ASV1472	Prevotellaceae	Prevotella	NA

ASV1473	Prevotellaceae	Prevotella_7	NA
ASV1474	NA	NA	NA
ASV1475	Corynebacteriaceae	Corynebacterium	NA
ASV1476	Moraxellaceae	Acinetobacter	bouvetii/haemolyticus/johnsonii/junii
ASV1477	Family_XI	Gemella	NA
ASV1478	Prevotellaceae	Prevotella_7	NA
ASV1479	Streptococcaceae	Streptococcus	NA
ASV1480	Neisseriaceae	Neisseria	NA
ASV1481	Porphyromonadaceae	Porphyromonas	NA
ASV1482	Porphyromonadaceae	Porphyromonas	NA
ASV1483	Prevotellaceae	Prevotella_7	NA
ASV1484	Mycoplasmataceae	Mycoplasma	spermatophilum
ASV1485	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1486	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1487	Pasteurellaceae	Aggregatibacter	NA
ASV1488	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1489	Prevotellaceae	Alloprevotella	NA
ASV1490	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1491	Prevotellaceae	Alloprevotella	NA
ASV1492	Pasteurellaceae	Haemophilus	NA
ASV1493	Enterobacteriaceae	NA	NA
ASV1494	Fusobacteriaceae	Fusobacterium	NA
ASV1495	NA	NA	NA
ASV1496	Neisseriaceae	Alysiella	NA
ASV1497	Prevotellaceae	Prevotella_2	NA
ASV1498	Lachnospiraceae	Johnsonella	NA

ASV1499	Burkholderiaceae	Delftia	acidovorans/cf./lacustris/tsuruhatensis
ASV1500	NA	NA	NA
ASV1501	Prevotellaceae	Prevotella_6	NA
ASV1502	Veillonellaceae	Selenomonas_3	NA
ASV1503	Prevotellaceae	Prevotella	NA
ASV1504	Leptotrichiaceae	Leptotrichia	NA
ASV1505	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1506	Weeksellaceae	Bergeyella	cardium
ASV1507	Prevotellaceae	Prevotella_7	NA
ASV1508	Fusobacteriaceae	Fusobacterium	NA
ASV1509	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1510	Leptotrichiaceae	Leptotrichia	NA
ASV1511	Prevotellaceae	Alloprevotella	NA
ASV1512	Atopobiaceae	Atopobium	NA
ASV1513	Fusobacteriaceae	Fusobacterium	NA
ASV1514	Fusobacteriaceae	Fusobacterium	NA
ASV1515	Prevotellaceae	Prevotella_7	NA
ASV1516	NA	NA	NA
ASV1517	Streptococcaceae	Streptococcus	NA
ASV1518	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1519	Mycoplasmataceae	Mycoplasma	NA
ASV1520	Fusobacteriaceae	Fusobacterium	NA
ASV1521	Prevotellaceae	Prevotella_7	NA
ASV1522	Prevotellaceae	Alloprevotella	NA
ASV1523	Fusobacteriaceae	Fusobacterium	NA
ASV1524	Leptotrichiaceae	Leptotrichia	NA

ASV1525	Corynebacteriaceae	Corynebacterium	NA
ASV1526	Weeksellaceae	Bergeyella	NA
ASV1527	Weeksellaceae	Bergeyella	NA
ASV1528	Fusobacteriaceae	Fusobacterium	NA
ASV1529	Prevotellaceae	Prevotella	NA
ASV1530	Streptococcaceae	Streptococcus	NA
ASV1531	Veillonellaceae	Selenomonas_3	NA
ASV1532	Prevotellaceae	Prevotella	NA
ASV1533	Desulfovibrionaceae	Desulfovibrio	NA
ASV1534	Leptotrichiaceae	Leptotrichia	NA
ASV1535	Prevotellaceae	Alloprevotella	NA
ASV1536	Fusobacteriaceae	Fusobacterium	NA
ASV1537	Neisseriaceae	Alysiella	NA
ASV1538	Lachnospiraceae	Roseburia	NA
ASV1539	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1540	Enterococcaceae	Vagococcus	NA
ASV1541	Spirochaetaceae	Treponema_2	NA
ASV1542	Carnobacteriaceae	Granulicatella	NA
ASV1543	Prevotellaceae	Alloprevotella	NA
ASV1544	Xanthobacteraceae	Bradyrhizobium	NA
ASV1545	NA	NA	NA
ASV1546	Streptococcaceae	Streptococcus	NA
ASV1547	Spirochaetaceae	Treponema_2	NA
ASV1548	Veillonellaceae	Veillonella	NA
ASV1549	Veillonellaceae	Veillonella	NA
ASV1550	Prevotellaceae	Prevotella_7	NA

ASV1551	Spirochaetaceae	Treponema_2	NA
ASV1552	Prevotellaceae	Prevotella_2	NA
ASV1553	Prevotellaceae	Prevotella	NA
ASV1554	Veillonellaceae	Dialister	NA
ASV1555	Neisseriaceae	Kingella	NA
ASV1556	Veillonellaceae	Veillonella	NA
ASV1557	Leptotrichiaceae	Leptotrichia	NA
ASV1558	Veillonellaceae	Veillonella	NA
ASV1559	Prevotellaceae	Prevotella_7	NA
ASV1560	Prevotellaceae	Prevotella	NA
ASV1561	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1562	Prevotellaceae	Alloprevotella	NA
ASV1563	Clostridiales_vadinBB60_group	NA	NA
ASV1564	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1565	Lachnospiraceae	Oribacterium	NA
ASV1566	NA	NA	NA
ASV1567	Prevotellaceae	Alloprevotella	NA
ASV1568	Veillonellaceae	NA	NA
ASV1569	Family_XIII	NA	NA
ASV1570	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1571	Erysipelotrichaceae	Erysipelotrichaceae_UCG-004	NA
ASV1572	Lachnospiraceae	Stomatobaculum	NA
ASV1573	Fusobacteriaceae	Fusobacterium	NA
ASV1574	Lachnospiraceae	Stomatobaculum	NA
ASV1575	Spirochaetaceae	Treponema_2	NA
ASV1576	Neisseriaceae	NA	NA

ASV1577	Mycoplasmataceae	Mycoplasma	NA
ASV1578	Streptococcaceae	Streptococcus	NA
ASV1579	Spirochaetaceae	Treponema_2	NA
ASV1580	Veillonellaceae	Selenomonas_3	NA
ASV1581	Prevotellaceae	Alloprevotella	NA
ASV1582	Prevotellaceae	Prevotella_2	NA
ASV1583	Prevotellaceae	Prevotella_7	NA
ASV1584	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1585	Carnobacteriaceae	Granulicatella	NA
ASV1586	Prevotellaceae	Prevotella	NA
ASV1587	Veillonellaceae	Veillonella	NA
ASV1588	Family_XIII	NA	NA
ASV1589	NA	NA	NA
ASV1590	Leptotrichiaceae	Leptotrichia	NA
ASV1591	Rs-M59_termite_group	NA	NA
ASV1592	Tannerellaceae	Tannerella	NA
ASV1593	NA	NA	NA
ASV1594	Family_XI	Gemella	NA
ASV1595	Lachnospiraceae	Stomatobaculum	NA
ASV1596	Streptococcaceae	Streptococcus	NA
ASV1597	Veillonellaceae	Veillonella	NA
ASV1598	Prevotellaceae	Prevotella_7	NA
ASV1599	Prevotellaceae	Prevotella	NA
ASV1600	NA	NA	NA
ASV1601	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1602	Carnobacteriaceae	Granulicatella	NA

ASV1603	Leptotrichiaceae	Leptotrichia	NA
ASV1604	Prevotellaceae	Prevotella	NA
ASV1605	Fusobacteriaceae	Fusobacterium	NA
ASV1606	Prevotellaceae	Prevotella_7	NA
ASV1607	Corynebacteriaceae	Corynebacterium_1	aurimucosum/jeikeium/propinquum/pseudodiphtheriticum
ASV1608	Prevotellaceae	Prevotella_2	NA
ASV1609	Porphyromonadaceae	Porphyromonas	NA
ASV1610	Rhodocyclaceae	Propionivibrio	NA
ASV1611	Prevotellaceae	Prevotella_2	NA
ASV1612	Prevotellaceae	Prevotella_6	NA
ASV1613	Family_XI	Parvimonas	NA
ASV1614	Prevotellaceae	Prevotella	NA
ASV1615	NA	NA	NA
ASV1616	Veillonellaceae	Veillonella	NA
ASV1617	Lachnospiraceae	Stomatobaculum	NA
ASV1618	NA	NA	NA
ASV1619	Carnobacteriaceae	Granulicatella	NA
ASV1620	Actinomycetaceae	Actinomyces	NA
ASV1621	Veillonellaceae	Selenomonas_3	NA
ASV1622	Prevotellaceae	Prevotella_6	NA
ASV1623	Prevotellaceae	Prevotella_7	NA
ASV1624	Leptotrichiaceae	Sneathia	amnii/sanguinegens
ASV1625	NA	NA	NA
ASV1626	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV1627	Prevotellaceae	Alloprevotella	NA
ASV1628	Family_XI	Gemella	NA

ASV1629	Peptococcaceae	Peptococcus	NA
ASV1630	Fusobacteriaceae	Fusobacterium	NA
ASV1631	Leptotrichiaceae	Leptotrichia	NA
ASV1632	NA	NA	NA
ASV1633	Fusobacteriaceae	Fusobacterium	NA
ASV1634	Prevotellaceae	Prevotella	NA
ASV1635	Micrococcaceae	Rothia	NA
ASV1636	Spirochaetaceae	Treponema_2	NA
ASV1637	Prevotellaceae	Alloprevotella	NA
ASV1638	Lactobacillaceae	Lactobacillus	kalixensis
ASV1639	Lachnospiraceae	Johnsonella	NA
ASV1640	Actinomycetaceae	Actinomyces	NA
ASV1641	Family_XIII	Mogibacterium	NA
ASV1642	Fusobacteriaceae	Fusobacterium	NA
ASV1643	Streptococcaceae	Streptococcus	NA
ASV1644	Leptotrichiaceae	Leptotrichia	NA
ASV1645	Prevotellaceae	Prevotella_6	NA
ASV1646	Prevotellaceae	Prevotella_7	NA
ASV1647	Fusobacteriaceae	Fusobacterium	NA
ASV1648	Prevotellaceae	Prevotella_7	NA
ASV1649	Prevotellaceae	Alloprevotella	NA
ASV1650	Atopobiaceae	Atopobium	NA
ASV1651	Carnobacteriaceae	Granulicatella	NA
ASV1652	Bifidobacteriaceae	Scardovia	NA
ASV1653	Veillonellaceae	Megasphaera	NA
ASV1654	Veillonellaceae	Selenomonas_3	NA

ASV1655	Leptotrichiaceae	Leptotrichia	NA
ASV1656	Lentimicrobiaceae	NA	NA
ASV1657	Flavobacteriaceae	Capnocytophaga	NA
ASV1658	Carnobacteriaceae	Granulicatella	NA
ASV1659	Streptococcaceae	Streptococcus	NA
ASV1660	Pasteurellaceae	Haemophilus	NA
ASV1661	Tannerellaceae	Tannerella	NA
ASV1662	Veillonellaceae	Selenomonas_3	NA
ASV1663	Spirochaetaceae	Treponema_2	NA
ASV1664	Prevotellaceae	Prevotella_7	NA
ASV1665	Flavobacteriaceae	Capnocytophaga	NA
ASV1666	Prevotellaceae	Prevotella_7	NA
ASV1667	Prevotellaceae	Alloprevotella	NA
ASV1668	Veillonellaceae	Selenomonas_3	NA
ASV1669	Synergistaceae	Fretibacterium	NA
ASV1670	Family_XIII	Mogibacterium	NA
ASV1671	Veillonellaceae	Dialister	NA
ASV1672	Neisseriaceae	Kingella	NA
ASV1673	Neisseriaceae	Neisseria	NA
ASV1674	Marinifilaceae	Odoribacter	NA
ASV1675	Moraxellaceae	Acinetobacter	baumannii/beijerinckii/bouvetii/calcoaceticus/dispersus/gyllenbergii/haemolyticus/junii/oryzae/parvus/tandoii/tjernbergiae
ASV1676	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1677	Spirochaetaceae	Treponema_2	NA
ASV1678	Spirochaetaceae	Treponema_2	NA
ASV1679	Actinomycetaceae	Actinomyces	NA
ASV1680	Spirochaetaceae	Treponema_2	NA

ASV1681	Veillonellaceae	Veillonella	NA
ASV1682	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1683	Family_XI	Peptoniphilus	lacrimalis
ASV1684	Veillonellaceae	Veillonella	NA
ASV1685	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1686	Fusobacteriaceae	Fusobacterium	NA
ASV1687	Fusobacteriaceae	Fusobacterium	NA
ASV1688	Pasteurellaceae	Haemophilus	NA
ASV1689	Prevotellaceae	Prevotella_7	NA
ASV1690	Mycoplasmataceae	Mycoplasma	NA
ASV1691	Veillonellaceae	Selenomonas_3	NA
ASV1692	Leptotrichiaceae	Leptotrichia	NA
ASV1693	Pasteurellaceae	Actinobacillus	NA
ASV1694	Streptococcaceae	Streptococcus	NA
ASV1695	Pasteurellaceae	Aggregatibacter	NA
ASV1696	Spirochaetaceae	Treponema_2	NA
ASV1697	Spirochaetaceae	Treponema_2	NA
ASV1698	Lachnospiraceae	Oribacterium	NA
ASV1699	Leptotrichiaceae	Leptotrichia	NA
ASV1700	Veillonellaceae	Veillonella	NA
ASV1701	Prevotellaceae	Prevotella_2	NA
ASV1702	Carnobacteriaceae	Granulicatella	NA
ASV1703	Streptococcaceae	Streptococcus	NA
ASV1704	Pseudomonadaceae	Pseudomonas	antarctica/auricularis/azotoformans/brenneri/canadensis/cedrina/cerasi/cf./costantinii/extremaustralis/extremorientalis/fluorescens/gessardii/grin
ASV1705	Family_XIII	NA	NA
ASV1706	Fusobacteriaceae	Fusobacterium	NA

ASV1707	NA	NA	NA
ASV1708	Spirochaetaceae	Treponema_2	NA
ASV1709	Fusobacteriaceae	Fusobacterium	NA
ASV1710	Prevotellaceae	Prevotella	NA
ASV1711	Flavobacteriaceae	Capnocytophaga	NA
ASV1712	Mycoplasmataceae	Mycoplasma	NA
ASV1713	Micrococcaceae	Rothia	NA
ASV1714	Streptococcaceae	Streptococcus	NA
ASV1715	Neisseriaceae	Neisseria	NA
ASV1716	Fusobacteriaceae	Fusobacterium	NA
ASV1717	Prevotellaceae	Prevotella_7	NA
ASV1718	Veillonellaceae	Veillonella	NA
ASV1719	Veillonellaceae	Veillonella	NA
ASV1720	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1721	Peptostreptococcaceae	Filifactor	NA
ASV1722	Porphyromonadaceae	Porphyromonas	NA
ASV1723	Weeksellaceae	Bergeyella	NA
ASV1724	Fusobacteriaceae	Fusobacterium	NA
ASV1725	Streptococcaceae	Streptococcus	NA
ASV1726	Leptotrichiaceae	Leptotrichia	NA
ASV1727	Carnobacteriaceae	Granulicatella	NA
ASV1728	Neisseriaceae	Alysiella	NA
ASV1729	Corynebacteriaceae	Corynebacterium_1	kroppenstedtii
ASV1730	NA	NA	NA
ASV1731	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV1732	Ruminococcaceae	Ruminococcaceae_UCG-014	NA

ASV1733	Campylobacteraceae	Campylobacter	NA
ASV1734	Prevotellaceae	Prevotella_7	NA
ASV1735	NA	NA	NA
ASV1736	Veillonellaceae	Veillonella	NA
ASV1737	Leptotrichiaceae	Leptotrichia	NA
ASV1738	Prevotellaceae	Prevotella_7	NA
ASV1739	Streptococcaceae	Streptococcus	NA
ASV1740	Mycoplasmataceae	Mycoplasma	NA
ASV1741	Streptococcaceae	Streptococcus	NA
ASV1742	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1743	Spirochaetaceae	Treponema_2	NA
ASV1744	Carnobacteriaceae	Granulicatella	NA
ASV1745	Lactobacillaceae	Lactobacillus	iners
ASV1746	Prevotellaceae	Alloprevotella	NA
ASV1747	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1748	Spirochaetaceae	Treponema_2	NA
ASV1749	Veillonellaceae	Megasphaera	NA
ASV1750	Prevotellaceae	Prevotella_7	NA
ASV1751	Lachnospiraceae	Johnsonella	NA
ASV1752	Cardiobacteriaceae	Cardiobacterium	NA
ASV1753	Lachnospiraceae	Butyrivibrio_2	NA
ASV1754	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1755	Spirochaetaceae	Treponema_2	NA
ASV1756	Streptococcaceae	Streptococcus	NA
ASV1757	Corynebacteriaceae	Corynebacterium	NA
ASV1758	Pasteurellaceae	NA	NA

ASV1759	Mycoplasmataceae	Mycoplasma	NA
ASV1760	Flavobacteriaceae	Capnocytophaga	NA
ASV1761	Veillonellaceae	Selenomonas_3	NA
ASV1762	Fusobacteriaceae	Fusobacterium	NA
ASV1763	Veillonellaceae	Veillonella	NA
ASV1764	Veillonellaceae	Selenomonas	sputigena
ASV1765	Leptotrichiaceae	Leptotrichia	NA
ASV1766	Prevotellaceae	Prevotella_7	NA
ASV1767	Lachnospiraceae	Oribacterium	NA
ASV1768	Prevotellaceae	Prevotella	NA
ASV1769	Paludibacteraceae	F0058	NA
ASV1770	Campylobacteraceae	Campylobacter	NA
ASV1771	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1772	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1773	Spirochaetaceae	Treponema_2	NA
ASV1774	Fusobacteriaceae	Fusobacterium	NA
ASV1775	Prevotellaceae	Alloprevotella	NA
ASV1776	Streptococcaceae	Lactococcus	garvieae/lactis/taiwanensis
ASV1777	Prevotellaceae	Prevotella	NA
ASV1778	Lachnospiraceae	NA	NA
ASV1779	Burkholderiaceae	NA	NA
ASV1780	Flavobacteriaceae	Capnocytophaga	NA
ASV1781	Family_XI	Gemella	NA
ASV1782	Flavobacteriaceae	Capnocytophaga	NA
ASV1783	Lachnospiraceae	NA	NA

ASV1784	Synergistaceae	Jonquetella	anthropi
ASV1785	Veillonellaceae	Veillonella	NA
ASV1786	NA	NA	NA
ASV1787	Lachnospiraceae	Johnsonella	NA
ASV1788	Veillonellaceae	Selenomonas_3	NA
ASV1789	Fusobacteriaceae	Fusobacterium	NA
ASV1790	Micrococcaceae	Rothia	NA
ASV1791	Paludibacteraceae	F0058	NA
ASV1792	Veillonellaceae	Veillonella	NA
ASV1793	Streptococcaceae	Streptococcus	NA
ASV1794	Micrococcaceae	Rothia	NA
ASV1795	Carnobacteriaceae	Granulicatella	NA
ASV1796	Veillonellaceae	NA	NA
ASV1797	Streptococcaceae	Streptococcus	NA
ASV1798	Carnobacteriaceae	Granulicatella	NA
ASV1799	Veillonellaceae	Megasphaera	NA
ASV1800	Campylobacteraceae	Campylobacter	NA
ASV1801	Moraxellaceae	Moraxella	NA
ASV1802	Flavobacteriaceae	Capnocytophaga	NA
ASV1803	Porphyromonadaceae	Porphyromonas	NA
ASV1804	Fusobacteriaceae	Fusobacterium	NA
ASV1805	Fusobacteriaceae	Fusobacterium	NA
ASV1806	Leptotrichiaceae	Leptotrichia	NA
ASV1807	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1808	Prevotellaceae	Prevotella_7	NA

ASV1809	Leptotrichiaceae	Leptotrichia	NA
ASV1810	Campylobacteraceae	Campylobacter	NA
ASV1811	Veillonellaceae	Veillonella	NA
ASV1812	Pasteurellaceae	Haemophilus	NA
ASV1813	Fusobacteriaceae	Fusobacterium	NA
ASV1814	Carnobacteriaceae	Granulicatella	NA
ASV1815	Prevotellaceae	Alloprevotella	NA
ASV1816	Prevotellaceae	Prevotella_7	NA
ASV1817	Fusobacteriaceae	Fusobacterium	NA
ASV1818	Streptococcaceae	Streptococcus	NA
ASV1819	Spirochaetaceae	Treponema_2	putidum
ASV1820	Burkholderiaceae	Lautropia	NA
ASV1821	Streptococcaceae	Streptococcus	NA
ASV1822	Spirochaetaceae	Treponema_2	NA
ASV1823	Leptotrichiaceae	Leptotrichia	NA
ASV1824	Tannerellaceae	Tannerella	NA
ASV1825	Leptotrichiaceae	Leptotrichia	NA
ASV1826	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1827	Neisseriaceae	Neisseria	NA
ASV1828	Carnobacteriaceae	Granulicatella	NA
ASV1829	Prevotellaceae	Prevotella_2	NA
ASV1830	Flavobacteriaceae	Capnocytophaga	NA
ASV1831	Veillonellaceae	Veillonella	NA
ASV1832	Flavobacteriaceae	Capnocytophaga	NA
ASV1833	Prevotellaceae	Alloprevotella	NA
ASV1834	Pasteurellaceae	Aggregatibacter	NA

ASV1835	Prevotellaceae	Prevotella	NA
ASV1836	Fusobacteriaceae	Fusobacterium	NA
ASV1837	NA	NA	NA
ASV1838	Streptococcaceae	Streptococcus	NA
ASV1839	Carnobacteriaceae	Granulicatella	NA
ASV1840	Weeksellaceae	Bergeyella	NA
ASV1841	Prevotellaceae	Prevotella_7	NA
ASV1842	Porphyromonadaceae	Porphyromonas	NA
ASV1843	Streptococcaceae	Streptococcus	NA
ASV1844	Prevotellaceae	Alloprevotella	NA
ASV1845	Carnobacteriaceae	Granulicatella	NA
ASV1846	Fusobacteriaceae	Fusobacterium	NA
ASV1847	Leptotrichiaceae	Leptotrichia	NA
ASV1848	Cardiobacteriaceae	Cardiobacterium	NA
ASV1849	Prevotellaceae	Prevotella_7	NA
ASV1850	Prevotellaceae	Prevotella	NA
ASV1851	Tannerellaceae	Tannerella	NA
ASV1852	Family_XI	Parvimonas	NA
ASV1853	Leptotrichiaceae	Leptotrichia	NA
ASV1854	Flavobacteriaceae	Capnocytophaga	NA
ASV1855	Prevotellaceae	Prevotella_7	NA
ASV1856	Mycoplasmataceae	Mycoplasma	NA
ASV1857	Halomonadaceae	Halomonas	aidingensis/alimentaria/cerina/daqiaonensis/desiderata/fontilapidosi/lutea/nitroreducens/rifensis/salina/stenophila/ventosae
ASV1858	Xanthomonadaceae	Stenotrophomonas	maltophilia
ASV1859	Streptococcaceae	Streptococcus	NA
ASV1860	Prevotellaceae	Prevotella_7	NA

ASV1861	Lachnospiraceae	Stomatobaculum	NA
ASV1862	Lactobacillaceae	Lactobacillus	NA
ASV1863	Lactobacillaceae	Lactobacillus	NA
ASV1864	Family_XI	Parvimonas	NA
ASV1865	Lachnospiraceae	Catonella	NA
ASV1866	Lachnospiraceae	Butyrivibrio_2	NA
ASV1867	Prevotellaceae	Alloprevotella	NA
ASV1868	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1869	Peptostreptococcaceae	Filifactor	NA
ASV1870	Pasteurellaceae	Mannheimia	NA
ASV1871	Fusobacteriaceae	Fusobacterium	NA
ASV1872	Muribaculaceae	NA	NA
ASV1873	Lactobacillaceae	Lactobacillus	delbrueckii/equicursoris
ASV1874	Moraxellaceae	Acinetobacter	baumannii/septicus/ursingii
ASV1875	Lactobacillaceae	Lactobacillus	fornicalis/jensenii
ASV1876	Veillonellaceae	Veillonella	NA
ASV1877	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1878	Veillonellaceae	Selenomonas_4	NA
ASV1879	Prevotellaceae	Prevotella	NA
ASV1880	Neisseriaceae	Simonsiella	NA
ASV1881	Fusobacteriaceae	Fusobacterium	NA
ASV1882	Lachnospiraceae	Oribacterium	NA
ASV1883	Streptococcaceae	Streptococcus	NA
ASV1884	NA	NA	NA
ASV1885	Streptococcaceae	Streptococcus	NA
ASV1886	Fusobacteriaceae	Fusobacterium	NA

ASV1887	Leptotrichiaceae	Leptotrichia	NA
ASV1888	Porphyromonadaceae	Porphyromonas	NA
ASV1889	Streptococcaceae	Streptococcus	NA
ASV1890	Mycoplasmataceae	Mycoplasma	NA
ASV1891	Pasteurellaceae	Haemophilus	NA
ASV1892	NA	NA	NA
ASV1893	Veillonellaceae	Megasphaera	NA
ASV1894	Family_XIII	NA	NA
ASV1895	Streptococcaceae	Streptococcus	NA
ASV1896	NA	NA	NA
ASV1897	NA	NA	NA
ASV1898	Spirochaetaceae	Treponema_2	NA
ASV1899	Fusobacteriaceae	Fusobacterium	NA
ASV1900	Prevotellaceae	Alloprevotella	NA
ASV1901	Lachnospiraceae	Stomatobaculum	NA
ASV1902	NA	NA	NA
ASV1903	Veillonellaceae	Veillonella	NA
ASV1904	Weeksellaceae	Bergeyella	NA
ASV1905	Lachnospiraceae	Oribacterium	NA
ASV1906	Prevotellaceae	Alloprevotella	NA
ASV1907	Actinomycetaceae	Actinomyces	NA
ASV1908	Peptostreptococcaceae	Peptostreptococcus	NA
ASV1909	Burkholderiaceae	Limnobacter	NA
ASV1910	Prevotellaceae	Prevotella_7	NA
ASV1911	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1912	Mycoplasmataceae	Mycoplasma	NA

ASV1913	Erysipelotrichaceae	Bulleidia	NA
ASV1914	Leptotrichiaceae	Leptotrichia	NA
ASV1915	Family_XI	Parvimonas	NA
ASV1916	Lachnospiraceae	Oribacterium	NA
ASV1917	Fusobacteriaceae	Fusobacterium	NA
ASV1918	Family_XI	Gemella	NA
ASV1919	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1920	Pasteurellaceae	Aggregatibacter	NA
ASV1921	Veillonellaceae	Veillonella	NA
ASV1922	Pasteurellaceae	Actinobacillus	NA
ASV1923	Spirochaetaceae	Treponema_2	NA
ASV1924	Lentimicrobiaceae	NA	NA
ASV1925	Fusobacteriaceae	Fusobacterium	NA
ASV1926	Veillonellaceae	Megasphaera	NA
ASV1927	Leptotrichiaceae	Leptotrichia	NA
ASV1928	Spirochaetaceae	Treponema_2	NA
ASV1929	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV1930	Prevotellaceae	Prevotella_7	NA
ASV1931	Corynebacteriaceae	Lawsonella	clevelandensis
ASV1932	Moraxellaceae	Enhydrobacter	aerosaccus
ASV1933	Streptococcaceae	Streptococcus	NA
ASV1934	Neisseriaceae	Neisseria	NA
ASV1935	Lachnospiraceae	Oribacterium	NA
ASV1936	NA	NA	NA
ASV1937	Lachnospiraceae	Stomatobaculum	NA
ASV1938	Prevotellaceae	Prevotella	NA

ASV1939	Prevotellaceae	Prevotella_7	NA
ASV1940	Spirochaetaceae	Treponema_2	NA
ASV1941	Atopobiaceae	Olsenella	NA
ASV1942	Lentimicrobiaceae	NA	NA
ASV1943	Fusobacteriaceae	Fusobacterium	NA
ASV1944	Campylobacteraceae	Campylobacter	NA
ASV1945	Lachnospiraceae	NA	NA
ASV1946	Lachnospiraceae	Stomatobaculum	NA
ASV1947	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1948	Veillonellaceae	Megasphaera	NA
ASV1949	Paludibacteraceae	F0058	NA
ASV1950	Fusobacteriaceae	Fusobacterium	NA
ASV1951	NA	NA	NA
ASV1952	NA	NA	NA
ASV1953	Streptococcaceae	Streptococcus	NA
ASV1954	Fusobacteriaceae	Fusobacterium	NA
ASV1955	Flavobacteriaceae	Capnocytophaga	NA
ASV1956	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV1957	NA	NA	NA
ASV1958	Streptococcaceae	Streptococcus	NA
ASV1959	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1960	Family_XI	Anaerococcus	NA
ASV1961	Enterobacteriaceae	Yersinia	NA
ASV1962	Leptotrichiaceae	Leptotrichia	NA
ASV1963	Spirochaetaceae	Treponema_2	NA
ASV1964	Family_XI	Gemella	NA

ASV1965	Desulfomicrobiaceae	Desulfomicrobium	NA
ASV1966	Leptotrichiaceae	Oceanivirga	NA
ASV1967	Carnobacteriaceae	Granulicatella	NA
ASV1968	Leptotrichiaceae	Leptotrichia	NA
ASV1969	Actinomycetaceae	Actinomyces	NA
ASV1970	NA	NA	NA
ASV1971	Prevotellaceae	Prevotella_7	NA
ASV1972	Prevotellaceae	Prevotella_2	NA
ASV1973	Leptotrichiaceae	Leptotrichia	NA
ASV1974	Carnobacteriaceae	Dolosigranulum	pigrum
ASV1975	Streptococcaceae	Streptococcus	NA
ASV1976	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV1977	Fusobacteriaceae	Fusobacterium	NA
ASV1978	Prevotellaceae	Prevotella_6	NA
ASV1979	Actinomycetaceae	Actinomyces	NA
ASV1980	Pasteurellaceae	Haemophilus	NA
ASV1981	Spirochaetaceae	Treponema_2	NA
ASV1982	Leptotrichiaceae	Leptotrichia	NA
ASV1983	Veillonellaceae	Veillonella	NA
ASV1984	Porphyromonadaceae	Porphyromonas	NA
ASV1985	NA	NA	NA
ASV1986	Veillonellaceae	NA	NA
ASV1987	Streptococcaceae	Streptococcus	oralis
ASV1988	Porphyromonadaceae	Porphyromonas	NA
ASV1989	NA	NA	NA
ASV1990	Carnobacteriaceae	Granulicatella	NA

ASV1991	Lactobacillaceae	Lactobacillus	NA
ASV1992	Fusobacteriaceae	Fusobacterium	NA
ASV1993	Fusobacteriaceae	Fusobacterium	NA
ASV1994	Fusobacteriaceae	Fusobacterium	NA
ASV1995	Flavobacteriaceae	Capnocytophaga	NA
ASV1996	Pseudomonadaceae	Pseudomonas	aeruginosa/alcaligenes/alcaliphila/asplenii/cf./chengduensis/composti/cuatrocienegasensis/entomophila/fluorescens/fulva/fuscovaginae
ASV1997	Prevotellaceae	Alloprevotella	NA
ASV1998	Fusobacteriaceae	Fusobacterium	NA
ASV1999	Prevotellaceae	Prevotella_7	NA
ASV2000	Spirochaetaceae	Treponema_2	NA
ASV2001	Veillonellaceae	Selenomonas_3	NA
ASV2002	Prevotellaceae	Prevotella_6	NA
ASV2003	Spirochaetaceae	Treponema_2	NA
ASV2004	Lachnospiraceae	Stomatobaculum	NA
ASV2005	Veillonellaceae	Dialister	NA
ASV2006	Porphyromonadaceae	Porphyromonas	NA
ASV2007	Prevotellaceae	Alloprevotella	NA
ASV2008	Lachnospiraceae	Stomatobaculum	NA
ASV2009	Lachnospiraceae	Oribacterium	NA
ASV2010	Leptotrichiaceae	Leptotrichia	NA
ASV2011	Prevotellaceae	Prevotella_7	NA
ASV2012	Carnobacteriaceae	Granulicatella	NA
ASV2013	Actinomycetaceae	Actinomyces	NA
ASV2014	Akkermansiaceae	Akkermansia	NA
ASV2015	Fusobacteriaceae	Fusobacterium	NA
ASV2016	Saccharimonadaceae	NA	NA

ASV2017	Lachnospiraceae	Oribacterium	NA
ASV2018	Family_XI	Gemella	NA
ASV2019	Veillonellaceae	Veillonella	NA
ASV2020	Prevotellaceae	Prevotella_7	NA
ASV2021	Burkholderiaceae	Comamonas	NA
ASV2022	Veillonellaceae	Selenomonas_3	NA
ASV2023	Lachnospiraceae	Johnsonella	NA
ASV2024	Leptotrichiaceae	Leptotrichia	NA
ASV2025	Atopobiaceae	Atopobium	NA
ASV2026	Prevotellaceae	Alloprevotella	NA
ASV2027	Prevotellaceae	NA	NA
ASV2028	Tannerellaceae	Tannerella	NA
ASV2029	Veillonellaceae	Veillonella	NA
ASV2030	Veillonellaceae	Selenomonas_3	NA
ASV2031	Tannerellaceae	Tannerella	NA
ASV2032	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2033	Spirochaetaceae	Treponema_2	NA
ASV2034	Lachnospiraceae	NA	NA
ASV2035	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2036	Pasteurellaceae	Rodentibacter	NA
ASV2037	Lachnospiraceae	Catonella	NA
ASV2038	Prevotellaceae	Prevotella_7	NA
ASV2039	Prevotellaceae	Prevotella_7	NA
ASV2040	Leptotrichiaceae	Leptotrichia	NA
ASV2041	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2042	Prevotellaceae	Alloprevotella	NA

ASV2043	NA	NA	NA
ASV2044	Pasteurellaceae	Actinobacillus	NA
ASV2045	Prevotellaceae	Prevotella_2	NA
ASV2046	Fusobacteriaceae	Fusobacterium	NA
ASV2047	Flavobacteriaceae	Capnocytophaga	NA
ASV2048	Fusobacteriaceae	Fusobacterium	NA
ASV2049	Lactobacillaceae	Lactobacillus	NA
ASV2050	Campylobacteraceae	Campylobacter	NA
ASV2051	NA	NA	NA
ASV2052	Actinomycetaceae	Actinomyces	NA
ASV2053	Lachnospiraceae	Catonella	NA
ASV2054	Fusobacteriaceae	Fusobacterium	NA
ASV2055	Fusobacteriaceae	Fusobacterium	NA
ASV2056	Prevotellaceae	Prevotella_6	NA
ASV2057	Flavobacteriaceae	Capnocytophaga	NA
ASV2058	Spirochaetaceae	Treponema_2	NA
ASV2059	Prevotellaceae	Alloprevotella	NA
ASV2060	Leptotrichiaceae	Leptotrichia	NA
ASV2061	Fusobacteriaceae	Fusobacterium	NA
ASV2062	Cardiobacteriaceae	Cardiobacterium	NA
ASV2063	Lachnospiraceae	Shuttleworthia	NA
ASV2064	Streptococcaceae	Streptococcus	NA
ASV2065	Prevotellaceae	Alloprevotella	NA
ASV2066	Spirochaetaceae	Treponema_2	parvum
ASV2067	Neisseriaceae	Kingella	NA
ASV2068	Campylobacteraceae	Campylobacter	NA

ASV2069	Prevotellaceae	Alloprevotella	NA
ASV2070	Lachnospiraceae	Stomatobaculum	NA
ASV2071	Lachnospiraceae	NA	NA
ASV2072	Carnobacteriaceae	Granulicatella	NA
ASV2073	Muribaculaceae	NA	NA
ASV2074	Atopobiaceae	Olsenella	NA
ASV2075	Leptotrichiaceae	Leptotrichia	NA
ASV2076	Streptococcaceae	Streptococcus	NA
ASV2077	Fusobacteriaceae	Fusobacterium	NA
ASV2078	Leptotrichiaceae	Leptotrichia	NA
ASV2079	Spirochaetaceae	Treponema_2	NA
ASV2080	Neisseriaceae	Neisseria	NA
ASV2081	Streptococcaceae	Streptococcus	NA
ASV2082	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2083	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2084	Moraxellaceae	Moraxella	catarrhalis/nonliquefaciens
ASV2085	Lactobacillaceae	Lactobacillus	casei/fermentum/helveticus
ASV2086	Prevotellaceae	Prevotella_7	NA
ASV2087	Leptotrichiaceae	Leptotrichia	NA
ASV2088	Prevotellaceae	Prevotella_7	NA
ASV2089	NA	NA	NA
ASV2090	Neisseriaceae	Kingella	NA
ASV2091	Lachnospiraceae	Stomatobaculum	NA
ASV2092	Lachnospiraceae	Oribacterium	NA
ASV2093	Leptotrichiaceae	Leptotrichia	NA
ASV2094	Leptotrichiaceae	Leptotrichia	NA

ASV2095	Carnobacteriaceae	Granulicatella	NA
ASV2096	Lachnospiraceae	Oribacterium	NA
ASV2097	Prevotellaceae	Prevotella_6	NA
ASV2098	Pasteurellaceae	Haemophilus	influenzae
ASV2099	Leuconostocaceae	Weissella	bombi/cibaria/confusa/hellenica/koreensis/paramesenteroides/soli
ASV2100	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2101	Carnobacteriaceae	Granulicatella	NA
ASV2102	Prevotellaceae	Prevotella_7	NA
ASV2103	Flavobacteriaceae	Capnocytophaga	NA
ASV2104	Lachnospiraceae	NA	NA
ASV2105	Fusobacteriaceae	Fusobacterium	NA
ASV2106	Lentimicrobiaceae	NA	NA
ASV2107	Streptococcaceae	Streptococcus	NA
ASV2108	Neisseriaceae	Eikenella	NA
ASV2109	Veillonellaceae	Selenomonas_4	NA
ASV2110	Spirochaetaceae	Treponema_2	NA
ASV2111	Fusobacteriaceae	Fusobacterium	NA
ASV2112	Prevotellaceae	Prevotella	NA
ASV2113	Atopobiaceae	Atopobium	NA
ASV2114	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2115	Porphyromonadaceae	Porphyromonas	NA
ASV2116	NA	NA	NA
ASV2117	Family_XIII	Mogibacterium	NA
ASV2118	Porphyromonadaceae	Porphyromonas	NA
ASV2119	Leptotrichiaceae	Leptotrichia	NA
ASV2120	Veillonellaceae	Veillonella	NA

ASV2121	Spirochaetaceae	Treponema_2	NA
ASV2122	Desulfomicrobiaceae	Desulfomicrobium	orale
ASV2123	Tannerellaceae	Tannerella	NA
ASV2124	Xanthomonadaceae	Xanthomonas	arboricola/axonopodis/bromi/campestris/cannabis/cassavae/citri/cucurbitae/cynarae/dyei/euvesicatoria/fragariae/gardneri/hortorum
ASV2125	Mycoplasmataceae	Mycoplasma	NA
ASV2126	Lachnospiraceae	Stomatobaculum	NA
ASV2127	Erysipelotrichaceae	Solobacterium	NA
ASV2128	Prevotellaceae	Prevotella	NA
ASV2129	Spirochaetaceae	Treponema_2	NA
ASV2130	Lachnospiraceae	Stomatobaculum	NA
ASV2131	Synergistaceae	Fretibacterium	NA
ASV2132	Family_XI	Gemella	NA
ASV2133	Porphyromonadaceae	Porphyromonas	NA
ASV2134	Weeksellaceae	Bergeyella	NA
ASV2135	Fusobacteriaceae	Fusobacterium	NA
ASV2136	F082	NA	NA
ASV2137	Flavobacteriaceae	Capnocytophaga	NA
ASV2138	Prevotellaceae	Alloprevotella	NA
ASV2139	Flavobacteriaceae	Capnocytophaga	NA
ASV2140	NA	NA	NA
ASV2141	Lentimicrobiaceae	NA	NA
ASV2142	Pasteurellaceae	NA	NA
ASV2143	Lachnospiraceae	Oribacterium	NA
ASV2144	Burkholderiaceae	Ralstonia	detusculanense/insidiosa/mannitolilytica/pickettii/solanacearum/syzygii
ASV2145	Veillonellaceae	Veillonella	NA
ASV2146	Carnobacteriaceae	Granulicatella	NA

ASV2147	Leptotrichiaceae	Leptotrichia	NA
ASV2148	Veillonellaceae	Dialister	NA
ASV2149	Atopobiaceae	Atopobium	NA
ASV2150	Spirochaetaceae	Treponema_2	NA
ASV2151	Leptotrichiaceae	Leptotrichia	NA
ASV2152	Carnobacteriaceae	Granulicatella	NA
ASV2153	Prevotellaceae	Prevotella_7	NA
ASV2154	Veillonellaceae	Selenomonas_3	NA
ASV2155	Family_XIII	NA	NA
ASV2156	NA	NA	NA
ASV2157	Streptococcaceae	Streptococcus	NA
ASV2158	Cardiobacteriaceae	Cardiobacterium	NA
ASV2159	NA	NA	NA
ASV2160	NA	NA	NA
ASV2161	Leptotrichiaceae	Leptotrichia	NA
ASV2162	Spirochaetaceae	Treponema_2	NA
ASV2163	Lachnospiraceae	Johnsonella	NA
ASV2164	Veillonellaceae	Veillonella	NA
ASV2165	Peptostreptococcaceae	Peptoanaerobacter	NA
ASV2166	Prevotellaceae	Prevotella_7	NA
ASV2167	Fusobacteriaceae	Fusobacterium	NA
ASV2168	Fusobacteriaceae	Fusobacterium	NA
ASV2169	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2170	Prevotellaceae	Prevotella_7	NA
ASV2171	Spirochaetaceae	Treponema_2	NA
ASV2172	Weeksellaceae	Bergeyella	NA

ASV2173	Tannerellaceae	Tannerella	NA
ASV2174	Lachnospiraceae	Catonella	NA
ASV2175	Veillonellaceae	Selenomonas_4	NA
ASV2176	Lachnospiraceae	Butyrivibrio_2	NA
ASV2177	Fusobacteriaceae	Fusobacterium	NA
ASV2178	Spirochaetaceae	Treponema_2	NA
ASV2179	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2180	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2181	Campylobacteraceae	Campylobacter	NA
ASV2182	Leptotrichiaceae	Leptotrichia	NA
ASV2183	Leptotrichiaceae	Leptotrichia	NA
ASV2184	Streptococcaceae	Streptococcus	equinus/gallolyticus/infantarius/lutetiensis/salivarius
ASV2185	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2186	Actinomycetaceae	F0332	NA
ASV2187	Veillonellaceae	Veillonella	NA
ASV2188	Leptotrichiaceae	Leptotrichia	NA
ASV2189	Veillonellaceae	Veillonella	NA
ASV2190	Spirochaetaceae	Treponema_2	NA
ASV2191	Lachnospiraceae	Oribacterium	NA
ASV2192	Prevotellaceae	Prevotella_7	NA
ASV2193	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2194	Streptococcaceae	Streptococcus	NA
ASV2195	Lachnospiraceae	Johnsonella	NA
ASV2196	Veillonellaceae	Selenomonas_3	NA
ASV2197	Mycoplasmataceae	Mycoplasma	NA
ASV2198	Ruminococcaceae	Ruminococcaceae_UCG-014	NA

ASV2199	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2200	Prevotellaceae	Alloprevotella	NA
ASV2201	Corynebacteriaceae	Corynebacterium	NA
ASV2202	Veillonellaceae	Selenomonas_3	NA
ASV2203	Paludibacteraceae	F0058	NA
ASV2204	Spirochaetaceae	Treponema_2	NA
ASV2205	Veillonellaceae	Dialister	NA
ASV2206	Flavobacteriaceae	Capnocytophaga	NA
ASV2207	NA	NA	NA
ASV2208	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2209	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2210	Veillonellaceae	Centipeda	NA
ASV2211	NA	NA	NA
ASV2212	Lachnospiraceae	Stomatobaculum	NA
ASV2213	Leptotrichiaceae	Leptotrichia	NA
ASV2214	Cardiobacteriaceae	NA	NA
ASV2215	NA	NA	NA
ASV2216	Lachnospiraceae	Stomatobaculum	NA
ASV2217	Prevotellaceae	Alloprevotella	NA
ASV2218	Family_XI	Parvimonas	NA
ASV2219	Fusobacteriaceae	Fusobacterium	NA
ASV2220	Lachnospiraceae	Oribacterium	NA
ASV2221	Lentimicrobiaceae	NA	NA
ASV2222	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2223	Fusobacteriaceae	Fusobacterium	NA
ASV2224	Lachnospiraceae	Lachnoanaerobaculum	NA

ASV2225	Prevotellaceae	Alloprevotella	NA
ASV2226	Fusobacteriaceae	Fusobacterium	NA
ASV2227	Neisseriaceae	Neisseria	NA
ASV2228	Veillonellaceae	Megasphaera	NA
ASV2229	Streptococcaceae	Streptococcus	NA
ASV2230	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2231	Desulfobulbaceae	Desulfobulbus	NA
ASV2232	Family_XIII	NA	NA
ASV2233	Erysipelotrichaceae	Solobacterium	NA
ASV2234	Family_XI	Gemella	NA
ASV2235	Weeksellaceae	Bergeyella	NA
ASV2236	Veillonellaceae	Veillonella	NA
ASV2237	Pasteurellaceae	Aggregatibacter	NA
ASV2238	Prevotellaceae	Prevotella	NA
ASV2239	Veillonellaceae	Veillonella	NA
ASV2240	Family_XIII	NA	NA
ASV2241	Fusobacteriaceae	Fusobacterium	NA
ASV2242	Mycoplasmataceae	Mycoplasma	NA
ASV2243	Porphyromonadaceae	Porphyromonas	NA
ASV2244	Prevotellaceae	Prevotella	NA
ASV2245	Peptostreptococcaceae	NA	NA
ASV2246	Akkermansiaceae	Akkermansia	muciniphila
ASV2247	Family_XIII	Family_XIII_UCG-001	NA
ASV2248	Veillonellaceae	Selenomonas_4	NA
ASV2249	Neisseriaceae	Neisseria	NA
ASV2250	Burkholderiaceae	Aquabacterium	NA

ASV2251	Fusobacteriaceae	Fusobacterium	NA
ASV2252	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2253	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2254	Spirochaetaceae	Treponema_2	NA
ASV2255	Lactobacillaceae	Lactobacillus	acidophilus/delbrueckii/fermentum/helveticus/rhamnosus
ASV2256	Aerococcaceae	Abiotrophia	NA
ASV2257	Spirochaetaceae	Treponema_2	NA
ASV2258	Veillonellaceae	Veillonella	NA
ASV2259	Veillonellaceae	NA	NA
ASV2260	Carnobacteriaceae	Granulicatella	NA
ASV2261	Spirochaetaceae	Treponema_2	NA
ASV2262	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2263	Lachnospiraceae	Oribacterium	NA
ASV2264	Family_XI	Gemella	NA
ASV2265	Spirochaetaceae	Treponema_2	NA
ASV2266	Mycoplasmataceae	Mycoplasma	NA
ASV2267	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2268	Spirochaetaceae	Treponema_2	NA
ASV2269	Burkholderiaceae	Lautropia	NA
ASV2270	Actinomycetaceae	Actinomyces	NA
ASV2271	Veillonellaceae	NA	NA
ASV2272	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2273	Mycoplasmataceae	Mycoplasma	NA
ASV2274	Leptotrichiaceae	Leptotrichia	NA
ASV2275	Prevotellaceae	Prevotella_9	NA
ASV2276	Lachnospiraceae	Johnsonella	NA

ASV2277	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2278	Burkholderiaceae	Sutterella	NA
ASV2279	Leptotrichiaceae	Leptotrichia	NA
ASV2280	Veillonellaceae	Veillonella	NA
ASV2281	Fusobacteriaceae	Fusobacterium	NA
ASV2282	Burkholderiaceae	Brachymonas	NA
ASV2283	Leptotrichiaceae	Leptotrichia	NA
ASV2284	Veillonellaceae	Veillonella	NA
ASV2285	Veillonellaceae	Selenomonas_3	NA
ASV2286	Carnobacteriaceae	Granulicatella	NA
ASV2287	Spirochaetaceae	Treponema_2	NA
ASV2288	Family_XI	W5053	NA
ASV2289	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2290	Pasteurellaceae	Actinobacillus	NA
ASV2291	Neisseriaceae	Alysiella	NA
ASV2292	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2293	Veillonellaceae	Veillonella	NA
ASV2294	Lachnospiraceae	Oribacterium	NA
ASV2295	Carnobacteriaceae	Granulicatella	NA
ASV2296	Neisseriaceae	Bergeriella	NA
ASV2297	Lachnospiraceae	NA	NA
ASV2298	Bifidobacteriaceae	Aeriscardovia	NA
ASV2299	Family_XI	Parvimonas	NA
ASV2300	Kineosporiaceae	Kineococcus	NA
ASV2301	Campylobacteraceae	Campylobacter	NA
ASV2302	Pasteurellaceae	Haemophilus	NA

ASV2303	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2304	Erysipelotrichaceae	Solobacterium	NA
ASV2305	Corynebacteriaceae	Corynebacterium	NA
ASV2306	Spirochaetaceae	Treponema_2	NA
ASV2307	Campylobacteraceae	Campylobacter	NA
ASV2308	Atopobiaceae	Olsenella	profusa
ASV2309	Veillonellaceae	Veillonella	NA
ASV2310	Flavobacteriaceae	Capnocytophaga	NA
ASV2311	Enterobacteriaceae	NA	NA
ASV2312	Family_XI	Gemella	NA
ASV2313	Veillonellaceae	Selenomonas	NA
ASV2314	Leptotrichiaceae	Leptotrichia	NA
ASV2315	Fusobacteriaceae	Fusobacterium	NA
ASV2316	Cardiobacteriaceae	Cardiobacterium	NA
ASV2317	Streptococcaceae	Streptococcus	NA
ASV2318	Family_XI	Gemella	NA
ASV2319	Fusobacteriaceae	Fusobacterium	NA
ASV2320	Leptotrichiaceae	Leptotrichia	NA
ASV2321	Neisseriaceae	NA	NA
ASV2322	Fusobacteriaceae	Fusobacterium	NA
ASV2323	Leptotrichiaceae	Leptotrichia	NA
ASV2324	Weeksellaceae	Bergeyella	NA
ASV2325	Actinomycetaceae	Actinomyces	NA
ASV2326	Prevotellaceae	Prevotella	NA
ASV2327	Streptococcaceae	Streptococcus	massiliensis
ASV2328	Campylobacteraceae	Campylobacter	NA

ASV2329	Fusobacteriaceae	Fusobacterium	NA
ASV2330	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2331	Carnobacteriaceae	Granulicatella	NA
ASV2332	Neisseriaceae	Neisseria	NA
ASV2333	Actinomycetaceae	Actinomyces	NA
ASV2334	Veillonellaceae	Veillonella	NA
ASV2335	Veillonellaceae	Selenomonas_4	NA
ASV2336	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2337	NA	NA	NA
ASV2338	Anaerolineaceae	Flexilinea	NA
ASV2339	Veillonellaceae	Dialister	propionicifaciens
ASV2340	Fusobacteriaceae	Fusobacterium	NA
ASV2341	Spirochaetaceae	Treponema_2	NA
ASV2342	Prevotellaceae	Prevotella	NA
ASV2343	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2344	Fusobacteriaceae	Fusobacterium	NA
ASV2345	Atopobiaceae	Atopobium	NA
ASV2346	Prevotellaceae	Prevotella_2	NA
ASV2347	Lactobacillaceae	Lactobacillus	rhamnosus/salivarius
ASV2348	Prevotellaceae	Prevotella	NA
ASV2349	Burkholderiaceae	NA	NA
ASV2350	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2351	Lachnospiraceae	Butyrivibrio_2	NA
ASV2352	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV2353	Family_XIII	Mogibacterium	NA
ASV2354	Veillonellaceae	Centipeda	NA

ASV2355	Lachnospiraceae	Stomatobaculum	NA
ASV2356	Lachnospiraceae	Stomatobaculum	NA
ASV2357	Prevotellaceae	Prevotella	disiens
ASV2358	Leptotrichiaceae	Leptotrichia	NA
ASV2359	Lachnospiraceae	Stomatobaculum	NA
ASV2360	Fusobacteriaceae	Fusobacterium	NA
ASV2361	Alcanivoracaceae	Alcanivorax	gelatiniphagus/marinus/venustensis
ASV2362	Neisseriaceae	Kingella	NA
ASV2363	Lactobacillaceae	Lactobacillus	pontis
ASV2364	NA	NA	NA
ASV2365	Beijerinckiaceae	Methylobacterium	aquaticum/platani
ASV2366	Fusobacteriaceae	Fusobacterium	NA
ASV2367	Lachnospiraceae	Shuttleworthia	NA
ASV2368	Tannerellaceae	Tannerella	NA
ASV2369	Micrococcaceae	Rothia	NA
ASV2370	Cardiobacteriaceae	Cardiobacterium	NA
ASV2371	Porphyromonadaceae	Porphyromonas	NA
ASV2372	Veillonellaceae	Selenomonas_3	NA
ASV2373	Family_XIII	NA	NA
ASV2374	Lachnospiraceae	Stomatobaculum	NA
ASV2375	Fusobacteriaceae	Fusobacterium	NA
ASV2376	Prevotellaceae	Prevotella_7	NA
ASV2377	Cardiobacteriaceae	Cardiobacterium	NA
ASV2378	Veillonellaceae	Selenomonas_4	NA
ASV2379	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2380	Spirochaetaceae	Treponema_2	NA

ASV2381	Prevotellaceae	Alloprevotella	NA
ASV2382	Leptotrichiaceae	Leptotrichia	NA
ASV2383	Lachnospiraceae	Oribacterium	NA
ASV2384	Desulfomicrobiaceae	Desulfoplanes	NA
ASV2385	Leptotrichiaceae	Leptotrichia	NA
ASV2386	Pseudomonadaceae	Pseudomonas	arsenicoxydans/avellanae/baetica/borealis/brassicacearum/clemancea/ficuserectae/fluorescens/frederiksbergensis/graminis
ASV2387	Erysipelotrichaceae	Solobacterium	NA
ASV2388	Neisseriaceae	Kingella	NA
ASV2389	Spirochaetaceae	Treponema_2	NA
ASV2390	Flavobacteriaceae	Capnocytophaga	NA
ASV2391	Spirochaetaceae	Treponema_2	NA
ASV2392	Fusobacteriaceae	Fusobacterium	NA
ASV2393	Prevotellaceae	Prevotella_7	NA
ASV2394	NA	NA	NA
ASV2395	NA	NA	NA
ASV2396	Lachnospiraceae	Stomatobaculum	NA
ASV2397	Veillonellaceae	Centipeda	NA
ASV2398	Spirochaetaceae	Treponema_2	NA
ASV2399	Tannerellaceae	Tannerella	NA
ASV2400	Carnobacteriaceae	Granulicatella	NA
ASV2401	Burkholderiaceae	Lautropia	NA
ASV2402	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2403	Family_XI	Parvimonas	NA
ASV2404	Leptotrichiaceae	Leptotrichia	NA
ASV2405	Prevotellaceae	Prevotella	NA
ASV2406	Leptotrichiaceae	Leptotrichia	NA

ASV2407	Micrococcaceae	Kocuria	dechangensis/flava/himachalensis/polaris/rosea/sediminis/turfanensis
ASV2408	Prevotellaceae	Prevotella_7	NA
ASV2409	Prevotellaceae	Alloprevotella	NA
ASV2410	Mitochondria	NA	NA
ASV2411	Prevotellaceae	Prevotella	NA
ASV2412	Leptotrichiaceae	Leptotrichia	NA
ASV2413	Family_XI	W5053	NA
ASV2414	Spirochaetaceae	Treponema_2	NA
ASV2415	Lachnospiraceae	Johnsonella	NA
ASV2416	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2417	Family_XIII	NA	NA
ASV2418	Tannerellaceae	Tannerella	NA
ASV2419	Weeksellaceae	Bergeyella	NA
ASV2420	Carnobacteriaceae	Granulicatella	NA
ASV2421	Veillonellaceae	Veillonella	NA
ASV2422	Lachnospiraceae	Butyrivibrio_2	NA
ASV2423	Prevotellaceae	Prevotella	NA
ASV2424	Prevotellaceae	Alloprevotella	NA
ASV2425	Weeksellaceae	Chryseobacterium	culicis/indologenes/joostei/lactis/rhizoplanae/sediminis/tructae
ASV2426	Campylobacteraceae	Campylobacter	NA
ASV2427	Porphyromonadaceae	Porphyromonas	NA
ASV2428	Prevotellaceae	Alloprevotella	NA
ASV2429	Prevotellaceae	Prevotella_7	NA
ASV2430	Spirochaetaceae	Treponema_2	NA
ASV2431	Prevotellaceae	Prevotella	NA
ASV2432	Prevotellaceae	Prevotella	NA

ASV2433	Prevotellaceae	Prevotella_2	NA
ASV2434	Caulobacteraceae	Brevundimonas	diminuta/naejangsanensis/vancanneytii
ASV2435	Spirochaetaceae	Treponema_2	NA
ASV2436	Lachnospiraceae	Stomatobaculum	NA
ASV2437	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV2438	Family_XIII	NA	NA
ASV2439	Prevotellaceae	Prevotella_7	NA
ASV2440	Fusobacteriaceae	Fusobacterium	NA
ASV2441	Streptococcaceae	Streptococcus	NA
ASV2442	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2443	Corynebacteriaceae	Corynebacterium	NA
ASV2444	Bifidobacteriaceae	Gardnerella	vaginalis
ASV2445	Fusobacteriaceae	Fusobacterium	NA
ASV2446	Prevotellaceae	Prevotella_7	NA
ASV2447	Prevotellaceae	Alloprevotella	NA
ASV2448	Burkholderiaceae	Lautropia	NA
ASV2449	Peptostreptococcaceae	Peptostreptococcus	NA
ASV2450	Lachnospiraceae	NA	NA
ASV2451	Flavobacteriaceae	Capnocytophaga	NA
ASV2452	Enterobacteriaceae	Pantoea	NA
ASV2453	Veillonellaceae	Dialister	NA
ASV2454	Veillonellaceae	Selenomonas_4	NA
ASV2455	Prevotellaceae	Prevotella_6	NA
ASV2456	Prevotellaceae	Alloprevotella	NA
ASV2457	Prevotellaceae	Prevotella	NA
ASV2458	Prevotellaceae	Prevotella_7	NA

ASV2459	Lachnospiraceae	Butyrivibrio_2	NA
ASV2460	Veillonellaceae	Selenomonas_4	NA
ASV2461	Veillonellaceae	Selenomonas	NA
ASV2462	Peptostreptococcaceae	Peptoanaerobacter	NA
ASV2463	Streptococcaceae	Streptococcus	NA
ASV2464	Lactobacillaceae	Lactobacillus	sanfranciscensis
ASV2465	Family_XIII	Mogibacterium	NA
ASV2466	Lachnospiraceae	Stomatobaculum	NA
ASV2467	Spirochaetaceae	Treponema_2	NA
ASV2468	Mycoplasmataceae	Mycoplasma	NA
ASV2469	Neisseriaceae	Kingella	NA
ASV2470	Lachnospiraceae	Stomatobaculum	NA
ASV2471	NA	NA	NA
ASV2472	Neisseriaceae	NA	NA
ASV2473	Veillonellaceae	Selenomonas_3	NA
ASV2474	Ruminococcaceae	Candidatus_Soleaferrea	NA
ASV2475	Burkholderiaceae	Brachymonas	NA
ASV2476	Erysipelotrichaceae	Bulleidia	NA
ASV2477	Carnobacteriaceae	Carnobacterium	gallinarum/inhibens/jeotgali/maltaromaticum/viridans
ASV2478	Desulfobulbaceae	Desulfobulbus	NA
ASV2479	Mycoplasmataceae	Mycoplasma	NA
ASV2480	Erysipelotrichaceae	Solobacterium	NA
ASV2481	Spirochaetaceae	Treponema_2	NA
ASV2482	Carnobacteriaceae	Granulicatella	NA
ASV2483	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2484	Mycoplasmataceae	Mycoplasma	NA

ASV2485	Fusobacteriaceae	Fusobacterium	NA
ASV2486	Synergistaceae	Jonquetella	anthropi
ASV2487	Peptococcaceae	Peptococcus	NA
ASV2488	NA	NA	NA
ASV2489	Moraxellaceae	Psychrobacter	halophilus/marincola/maritimus/psychrophilus/submarinus
ASV2490	Pseudomonadaceae	Pseudomonas	antarctica/brenneri/chlororaphis/denitrificans/extremaustralis/fluorescens/gessardii/marginalis/meridiana/palleroniana/panacis/poae
ASV2491	NA	NA	NA
ASV2492	Desulfovibrionaceae	Desulfovibrio	NA
ASV2493	Lachnospiraceae	Butyrivibrio_2	NA
ASV2494	Veillonellaceae	NA	NA
ASV2495	Rhizobiaceae	Brucella	NA
ASV2496	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2497	Spirochaetaceae	Treponema_2	NA
ASV2498	Actinomycetaceae	Actinomyces	NA
ASV2499	Lachnospiraceae	Catonella	NA
ASV2500	Neisseriaceae	Kingella	NA
ASV2501	Peptostreptococcaceae	Peptoanaerobacter	NA
ASV2502	Flavobacteriaceae	Capnocytophaga	NA
ASV2503	NA	NA	NA
ASV2504	Porphyromonadaceae	Porphyromonas	NA
ASV2505	Actinomycetaceae	Actinomyces	NA
ASV2506	Ruminococcaceae	NA	NA
ASV2507	Carnobacteriaceae	Granulicatella	NA
ASV2508	Prevotellaceae	Prevotella_7	NA
ASV2509	Leptotrichiaceae	Leptotrichia	NA
ASV2510	Leptotrichiaceae	Leptotrichia	NA

ASV2511	Neisseriaceae	Neisseria	lactamica/meningitidis/polysaccharea
ASV2512	Spirochaetaceae	Treponema_2	NA
ASV2513	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2514	Fusobacteriaceae	Fusobacterium	NA
ASV2515	Burkholderiaceae	Lautropia	NA
ASV2516	Leptotrichiaceae	Leptotrichia	NA
ASV2517	Carnobacteriaceae	Granulicatella	NA
ASV2518	Prevotellaceae	Prevotella	NA
ASV2519	Spirochaetaceae	Treponema_2	NA
ASV2520	Lachnospiraceae	Oribacterium	NA
ASV2521	Leptotrichiaceae	Leptotrichia	NA
ASV2522	Neisseriaceae	NA	NA
ASV2523	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2524	Prevotellaceae	Prevotella_2	NA
ASV2525	Flavobacteriaceae	Capnocytophaga	NA
ASV2526	Family_XIII	NA	NA
ASV2527	Prevotellaceae	Prevotella	NA
ASV2528	Atopobiaceae	Atopobium	vaginae
ASV2529	Prevotellaceae	Prevotella_2	NA
ASV2530	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2531	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2532	Leptotrichiaceae	Leptotrichia	NA
ASV2533	Leptotrichiaceae	Leptotrichia	NA
ASV2534	Moraxellaceae	Acinetobacter	idrijaensis/lwoffii
ASV2535	Campylobacteraceae	Campylobacter	ureolyticus
ASV2536	Porphyromonadaceae	Porphyromonas	NA

ASV2537	Lachnospiraceae	Johnsonella	NA
ASV2538	Leptotrichiaceae	Leptotrichia	NA
ASV2539	Family_XI	Parvimonas	NA
ASV2540	Psychromonadaceae	Psychromonas	NA
ASV2541	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV2542	Spirochaetaceae	Treponema_2	NA
ASV2543	Spirochaetaceae	Treponema_2	NA
ASV2544	Mycoplasmataceae	Ureaplasma	parvum/urealyticum
ASV2545	Spirochaetaceae	Treponema_2	NA
ASV2546	Prevotellaceae	Prevotella_7	NA
ASV2547	NA	NA	NA
ASV2548	Campylobacteraceae	Campylobacter	NA
ASV2549	Lachnospiraceae	Stomatobaculum	NA
ASV2550	Carnobacteriaceae	Granulicatella	NA
ASV2551	Veillonellaceae	Dialister	NA
ASV2552	Lachnospiraceae	Shuttleworthia	NA
ASV2553	Family_XIII	Mogibacterium	NA
ASV2554	Family_XI	Parvimonas	NA
ASV2555	Cardiobacteriaceae	Cardiobacterium	NA
ASV2556	Neisseriaceae	Kingella	NA
ASV2557	Pseudomonadaceae	Pseudomonas	monteilii/plecoglossicida/putida/taiwanensis
ASV2558	Veillonellaceae	Selenomonas_4	NA
ASV2559	Leptotrichiaceae	Leptotrichia	NA
ASV2560	Lachnospiraceae	Fusicatenibacter	saccharivorans
ASV2561	Leptotrichiaceae	Leptotrichia	NA
ASV2562	Prevotellaceae	Prevotella	NA

ASV2563	Aerococcaceae	Abiotrophia	NA
ASV2564	Neisseriaceae	Kingella	NA
ASV2565	Prevotellaceae	Alloprevotella	NA
ASV2566	Prevotellaceae	Alloprevotella	NA
ASV2567	Caulobacteraceae	NA	NA
ASV2568	Flavobacteriaceae	Capnocytophaga	NA
ASV2569	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2570	Bifidobacteriaceae	Bifidobacterium	catenulatum/gallicum/kashiwanohense/pseudocatenulatum
ASV2571	Spirochaetaceae	Treponema_2	NA
ASV2572	Prevotellaceae	Prevotella	NA
ASV2573	Fusobacteriaceae	Fusobacterium	NA
ASV2574	Lachnospiraceae	Oribacterium	NA
ASV2575	Lachnospiraceae	Johnsonella	NA
ASV2576	Lachnospiraceae	Oribacterium	NA
ASV2577	Neisseriaceae	Kingella	kingae
ASV2578	Veillonellaceae	Selenomonas_4	NA
ASV2579	NA	NA	NA
ASV2580	Veillonellaceae	Selenomonas_4	NA
ASV2581	Enterococcaceae	Enterococcus	NA
ASV2582	Lachnospiraceae	Catonella	NA
ASV2583	NA	NA	NA
ASV2584	Fusobacteriaceae	Fusobacterium	NA
ASV2585	Spirochaetaceae	Treponema_2	NA
ASV2586	Prevotellaceae	Prevotella	NA
ASV2587	Prevotellaceae	Prevotella_7	NA
ASV2588	Lentimicrobiaceae	NA	NA

ASV2589	Aeromonadaceae	Aeromonas	allosaccharophila/aquatica/aquatilis/australiensis/bestiarum/bivalvium/cavernicola/caviae/crassostreae/dhakensis/diversa/encheleia
ASV2590	Spirochaetaceae	Treponema_2	NA
ASV2591	Spirochaetaceae	Treponema_2	NA
ASV2592	Mycoplasmataceae	Mycoplasma	NA
ASV2593	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2594	Lachnospiraceae	NA	NA
ASV2595	Weeksellaceae	Chryseobacterium	arthrosphaerae/contaminans/cucumeris/culicis/gleum/indologenes/jejuense/joostei/lactis/letacus/oncorhynchi/rhizosphaerae/tructae
ASV2596	Neisseriaceae	NA	NA
ASV2597	Leptotrichiaceae	Leptotrichia	NA
ASV2598	Leptotrichiaceae	Leptotrichia	NA
ASV2599	Spirochaetaceae	Treponema_2	NA
ASV2600	Atopobiaceae	Atopobium	NA
ASV2601	Actinomycetaceae	Actinomyces	NA
ASV2602	Lentimicrobiaceae	NA	NA
ASV2603	Lachnospiraceae	Lachnoclostridium	NA
ASV2604	Leptotrichiaceae	Leptotrichia	NA
ASV2605	Bifidobacteriaceae	Gardnerella	vaginalis
ASV2606	Neisseriaceae	Neisseria	NA
ASV2607	Veillonellaceae	NA	NA
ASV2608	Prevotellaceae	Prevotella_6	NA
ASV2609	Spirochaetaceae	Treponema_2	NA
ASV2610	Lachnospiraceae	Catonella	NA
ASV2611	Veillonellaceae	Selenomonas_3	NA
ASV2612	Lachnospiraceae	Catonella	NA
ASV2613	Campylobacteraceae	Campylobacter	NA
ASV2614	Beijerinckiaceae	Methylobacterium	radiotolerans

ASV2615	Porphyromonadaceae	Porphyromonas	NA
ASV2616	Veillonellaceae	Veillonella	NA
ASV2617	Mycoplasmataceae	Mycoplasma	NA
ASV2618	Mycoplasmataceae	Mycoplasma	NA
ASV2619	Spirochaetaceae	Treponema_2	NA
ASV2620	Prevotellaceae	Prevotella	NA
ASV2621	Spirochaetaceae	Treponema_2	NA
ASV2622	Desulfovibrionaceae	Desulfovibrio	NA
ASV2623	Lachnospiraceae	Oribacterium	NA
ASV2624	Rhodobacteraceae	Paracoccus	aminophilus/aminovorans/caeni/chinensis/communis/denitrificans/halophilus/huijuniae
ASV2625	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2626	Prevotellaceae	Prevotella_2	NA
ASV2627	NA	NA	NA
ASV2628	Spirochaetaceae	Treponema_2	NA
ASV2629	Spirochaetaceae	Treponema_2	NA
ASV2630	Prevotellaceae	Prevotella	NA
ASV2631	Flavobacteriaceae	Capnocytophaga	NA
ASV2632	Moraxellaceae	Psychrobacter	cibarius/fozii/glacincola/immobilis/namhaensis
ASV2633	Prevotellaceae	Prevotella_6	NA
ASV2634	Bifidobacteriaceae	Bifidobacterium	NA
ASV2635	Leptotrichiaceae	Leptotrichia	NA
ASV2636	Pasteurellaceae	Actinobacillus	NA
ASV2637	Spirochaetaceae	Treponema_2	NA
ASV2638	Lachnospiraceae	Catonella	NA
ASV2639	Streptococcaceae	Streptococcus	NA
ASV2640	Saccharimonadaceae	NA	NA

ASV2641	Spirochaetaceae	Treponema_2	NA
ASV2642	Lachnospiraceae	Johnsonella	NA
ASV2643	Family_XIII	NA	NA
ASV2644	Actinomycetaceae	F0332	NA
ASV2645	Geodermatophilaceae	Blastococcus	aggregatus/colisei/saxobsidens
ASV2646	Neisseriaceae	Alysiella	NA
ASV2647	Lachnospiraceae	Stomatobaculum	NA
ASV2648	Leptotrichiaceae	Streptobacillus	NA
ASV2649	Veillonellaceae	Veillonella	NA
ASV2650	Fusobacteriaceae	Fusobacterium	NA
ASV2651	Weeksellaceae	Bergeyella	NA
ASV2652	p-2534-18B5_gut_group	NA	NA
ASV2653	Prevotellaceae	Prevotella	NA
ASV2654	Enterobacteriaceae	Escherichia/Shigella	NA
ASV2655	Lachnospiraceae	Lachnospiraceae_NK4A136_group	NA
ASV2656	Prevotellaceae	Prevotellaceae_Ga6A1_group	NA
ASV2657	Leptotrichiaceae	Leptotrichia	NA
ASV2658	Enterobacteriaceae	Serratia	NA
ASV2659	Lachnospiraceae	Moryella	NA
ASV2660	Lachnospiraceae	Stomatobaculum	NA
ASV2661	Campylobacteraceae	Campylobacter	NA
ASV2662	Saccharimonadaceae	NA	NA
ASV2663	Moraxellaceae	Psychrobacter	arcticus/cryohalolentis/fozii/frigidicola/glacialis/glacincola/immobilis/luti/okhotskensis
ASV2664	Pasteurellaceae	Mannheimia	NA
ASV2665	Ruminococcaceae	Ruminococcaceae_UCG-014	NA

ASV2666	Ruminococcaceae	Ruminiclostridium_5	NA
ASV2667	Rhizobiaceae	Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium	NA
ASV2668	Carnobacteriaceae	Granulicatella	NA
ASV2669	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2670	NA	NA	NA
ASV2671	Atopobiaceae	Atopobium	NA
ASV2672	Veillonellaceae	Selenomonas	NA
ASV2673	Actinomycetaceae	Actinomyces	timonensis
ASV2674	Lachnospiraceae	NA	NA
ASV2675	Lachnospiraceae	Johnsonella	NA
ASV2676	Mitochondria	NA	NA
ASV2677	NA	NA	NA
ASV2678	Pseudomonadaceae	Pseudomonas	alcaliphila/mendocina/oleovorans/putida/songnenensis/stutzeri/xanthomarina
ASV2679	NA	NA	NA
ASV2680	JG30-KF-AS9	NA	NA
ASV2681	Lachnospiraceae	Oribacterium	NA
ASV2682	Micromonosporaceae	NA	NA
ASV2683	Pseudomonadaceae	Pseudomonas	alcaligenes
ASV2684	Veillonellaceae	Selenomonas	NA
ASV2685	Akkermansiaceae	Akkermansia	muciniphila
ASV2686	Campylobacteraceae	Campylobacter	NA
ASV2687	Veillonellaceae	Dialister	NA
ASV2688	Bacillaceae	NA	NA
ASV2689	NA	NA	NA
ASV2690	Mycoplasmataceae	Mycoplasma	NA

ASV2691	Prevotellaceae	Prevotella	timonensis
ASV2692	Mycobacteriaceae	Mycobacterium	asiaticum/gordonae/paragordonae
ASV2693	Prevotellaceae	Prevotella_7	NA
ASV2694	Campylobacteraceae	Campylobacter	NA
ASV2695	Family_XIII	NA	NA
ASV2696	Spirochaetaceae	Treponema_2	NA
ASV2697	Fusobacteriaceae	Fusobacterium	equinum/gonidiaformans
ASV2698	Weeksellaceae	Empedobacter	brevis
ASV2699	Prevotellaceae	Prevotella	NA
ASV2700	Family_XIII	NA	NA
ASV2701	Family_XI	Parvimonas	NA
ASV2702	Lachnospiraceae	Johnsonella	NA
ASV2703	Bacteroidaceae	Bacteroides	NA
ASV2704	Peptostreptococcaceae	Filifactor	NA
ASV2705	Rhodocyclaceae	Propionivibrio	NA
ASV2706	Prevotellaceae	Alloprevotella	NA
ASV2707	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV2708	Family_XI	Peptoniphilus	asaccharolyticus/grossensis/harei
ASV2709	Porphyromonadaceae	Porphyromonas	NA
ASV2710	Caulobacteraceae	Brevundimonas	bullata/halotolerans
ASV2711	Burkholderiaceae	NA	NA
ASV2712	Pasteurellaceae	Mannheimia	NA
ASV2713	Saccharimonadaceae	NA	NA
ASV2714	Moraxellaceae	Psychrobacter	NA
ASV2715	Family_XI	Parvimonas	NA
ASV2716	Prevotellaceae	Prevotella_2	NA

ASV2717	Lachnospiraceae	NA	NA
ASV2718	Lachnospiraceae	Agathobacter	NA
ASV2719	Cardiobacteriaceae	NA	NA
ASV2720	Acholeplasmataceae	Acholeplasma	NA
ASV2721	Veillonellaceae	Selenomonas_3	NA
ASV2722	Christensenellaceae	Christensenellaceae_R-7_group	NA
ASV2723	Flavobacteriaceae	Capnocytophaga	NA
ASV2724	Spirochaetaceae	Treponema_2	NA
ASV2725	Lachnospiraceae	Oribacterium	NA
ASV2726	NA	NA	NA
ASV2727	Mitochondria	NA	NA
ASV2728	Flavobacteriaceae	Capnocytophaga	NA
ASV2729	Leptotrichiaceae	Leptotrichia	NA
ASV2730	Lachnospiraceae	Johnsonella	NA
ASV2731	Leptotrichiaceae	Leptotrichia	NA
ASV2732	Lachnospiraceae	Moryella	indoligenes
ASV2733	Spirochaetaceae	Treponema_2	NA
ASV2734	Lachnospiraceae	Butyrivibrio_2	NA
ASV2735	Carnobacteriaceae	Granulicatella	NA
ASV2736	Cardiobacteriaceae	Cardiobacterium	NA
ASV2737	Burkholderiaceae	Ralstonia	insidiosa/pickettii/solanacearum/syzygii
ASV2738	Prevotellaceae	Prevotella_7	NA
ASV2739	Xanthomonadaceae	Stenotrophomonas	maltophilia/rhizophila
ASV2740	Lachnospiraceae	Oribacterium	NA
ASV2741	Family_XIII	Mogibacterium	NA
ASV2742	Prevotellaceae	Prevotella_9	copri

ASV2743	0319-6G20	NA	NA
ASV2744	Veillonellaceae	Selenomonas	NA
ASV2745	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2746	NA	NA	NA
ASV2747	Rikenellaceae	NA	NA
ASV2748	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia	NA
ASV2749	Pseudomonadaceae	Pseudomonas	aeruginosa/alkylphenolica/argentinensis/benzenivorans/cremoricolorata/cuatrocienegasensis/flavescens/fluorescens/fulva/glareae
ASV2750	Lachnospiraceae	Oribacterium	NA
ASV2751	Lachnospiraceae	Roseburia	NA
ASV2752	Micrococcaceae	Renibacterium	NA
ASV2753	Erysipelotrichaceae	Bulleidia	NA
ASV2754	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV2755	Tannerellaceae	Tannerella	NA
ASV2756	NA	NA	NA
ASV2757	Leptotrichiaceae	Leptotrichia	NA
ASV2758	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2759	Veillonellaceae	Veillonella	NA
ASV2760	Spirochaetaceae	Treponema_2	NA
ASV2761	Bifidobacteriaceae	Parascardovia	NA
ASV2762	Lachnospiraceae	Johnsonella	NA
ASV2763	Veillonellaceae	Selenomonas_3	NA
ASV2764	Burkholderiaceae	Schlegelella	aquatica/thermodepolymerans
ASV2765	Flavobacteriaceae	Capnocytophaga	NA
ASV2766	Prevotellaceae	Prevotella	NA
ASV2767	Lachnospiraceae	Oribacterium	NA

ASV2768	Lactobacillaceae	Lactobacillus	gastricus
ASV2769	Ruminococcaceae	Ruminococcus_2	bromii
ASV2770	Clostridiales_vadinBB60_group	NA	NA
ASV2771	Xanthomonadaceae	Thermomonas	NA
ASV2772	Streptococcaceae	Streptococcus	agalactiae/pyogenes
ASV2773	Geodermatophilaceae	Modestobacter	caceresii/marinus/multiseptatus/versicolor
ASV2774	Spirosomaceae	Dyadobacter	NA
ASV2775	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV2776	Actinomycetaceae	Actinomyces	NA
ASV2777	Polyangiaceae	Pajaroellobacter	NA
ASV2778	Porphyromonadaceae	Porphyromonas	NA
ASV2779	Porphyromonadaceae	Porphyromonas	NA
ASV2780	NA	NA	NA
ASV2781	Lachnospiraceae	NA	NA
ASV2782	Lachnospiraceae	Johnsonella	NA
ASV2783	Atopobiaceae	Atopobium	NA
ASV2784	Streptococcaceae	Streptococcus	NA
ASV2785	Family_XI	Parvimonas	NA
ASV2786	Lactobacillaceae	Lactobacillus	NA
ASV2787	Spirochaetaceae	Treponema_2	NA
ASV2788	Marinifilaceae	Odoribacter	NA
ASV2789	Lachnospiraceae	Shuttleworthia	NA
ASV2790	Rikenellaceae	Alistipes	NA
ASV2791	Sphingobacteriaceae	Sphingobacterium	spiritivorum
ASV2792	Acholeplasmataceae	Acholeplasma	NA
ASV2793	Sphingomonadaceae	Sphingopyxis	alaskensis/bauzanensis/chilensis/fribergensis/ginsengisoli/italica/panaciterrae

ASV2794	Mycoplasmataceae	Mycoplasma	NA
ASV2795	Moraxellaceae	Acinetobacter	baumannii/calcoaceticus/indicus/radioresistens/refrigeratoris/variabilis
ASV2796	Moraxellaceae	Acinetobacter	bouvetii/calcoaceticus/haemolyticus/johnsonii/schindleri
ASV2797	Erysipelotrichaceae	Eggerthia	NA
ASV2798	Pseudonocardiaceae	NA	NA
ASV2799	Mycoplasmataceae	Mycoplasma	NA
ASV2800	Veillonellaceae	NA	NA
ASV2801	Burkholderiaceae	NA	NA
ASV2802	Pseudomonadaceae	Pseudomonas	brassicacearum/chlororaphis/corrugata/fluorescens/jessenii/kilonensis/mohnii/moorei/putida/umsongensis/vancouverensis
ASV2803	Mitochondria	NA	NA
ASV2804	Lentimicrobiaceae	NA	NA
ASV2805	Tannerellaceae	Tannerella	NA
ASV2806	Streptococcaceae	Streptococcus	NA
ASV2807	Lachnospiraceae	Blautia	NA
ASV2808	Peptostreptococcaceae	Intestinibacter	bartlettii
ASV2809	Actinomycetaceae	Actinomyces	NA
ASV2810	Saccharimonadaceae	NA	NA
ASV2811	Prevotellaceae	Prevotella	NA
ASV2812	Lachnospiraceae	Johnsonella	NA
ASV2813	Caulobacteraceae	Brevundimonas	albigilva/nasdae/vesicularis
ASV2814	Enterobacteriaceae	Escherichia/Shigella	NA
ASV2815	Lachnospiraceae	Johnsonella	NA
ASV2816	Rubrobacteriaceae	Rubrobacter	NA
ASV2817	Weeksellaceae	Cloacibacterium	normanense/rupense
ASV2818	Lentimicrobiaceae	NA	NA
ASV2819	Hyphomicrobiaceae	Hyphomicrobium	NA

ASV2820	Prevotellaceae	Prevotella_7	NA
ASV2821	Mitochondria	NA	NA
ASV2822	Burkholderiaceae	NA	NA
ASV2823	Family_XI	W5053	NA
ASV2824	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV2825	Porphyromonadaceae	Porphyromonas	NA
ASV2826	Leptotrichiaceae	Leptotrichia	NA
ASV2827	Bacteriovoracaceae	Peredibacter	NA
ASV2828	Xanthomonadaceae	Stenotrophomonas	maltophilia
ASV2829	Lactobacillaceae	Lactobacillus	NA
ASV2830	Sphingobacteriaceae	Pedobacter	agri/alluvionis/borealis/ginsenosidimutans/humicola/jeongneungensis/kyungheensis/roseus/sandarakinus/soli/suwonensis/terrae
ASV2831	Fusobacteriaceae	Fusobacterium	NA
ASV2832	Family_XI	Parvimonas	NA
ASV2833	Flavobacteriaceae	Capnocytophaga	NA
ASV2834	Veillonellaceae	Selenomonas	NA
ASV2835	Micrococcaceae	Kocuria	NA
ASV2836	Lentimicrobiaceae	NA	NA
ASV2837	Spirochaetaceae	Treponema_2	NA
ASV2838	Veillonellaceae	Selenomonas_3	NA
ASV2839	Lactobacillaceae	Lactobacillus	NA
ASV2840	Bifidobacteriaceae	Bifidobacterium	NA
ASV2841	Lachnospiraceae	Blautia	NA
ASV2842	Micrococcaceae	Micrococcus	alkanovora/aloeverae/antarcticus/cohnii/endophyticus/flavus/indicus/luteus/lylae/thailandicus/yunnanensis
ASV2843	Chitinophagaceae	NA	NA
ASV2844	Mycoplasmataceae	Mycoplasma	NA

ASV2845	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2846	Neisseriaceae	Eikenella	NA
ASV2847	Burkholderiaceae	Paucibacter	NA
ASV2848	Family_XIII	Mogibacterium	NA
ASV2849	Burkholderiaceae	Comamonas	NA
ASV2850	Lachnospiraceae	Oribacterium	NA
ASV2851	Burkholderiaceae	Noviherbaspirillum	NA
ASV2852	Lachnospiraceae	Shuttleworthia	NA
ASV2853	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2854	Bifidobacteriaceae	Bifidobacterium	animalis
ASV2855	Weeksellaceae	Chryseobacterium	balustinum/indoltheticum/piscicola/piscium/scophthalmum/soldanellicola
ASV2856	Weeksellaceae	Bergeyella	NA
ASV2857	Mitochondria	NA	NA
ASV2858	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2859	Prevotellaceae	Prevotella_7	NA
ASV2860	Enterobacteriaceae	Escherichia/Shigella	NA
ASV2861	Alcanivoracaceae	Alcanivorax	NA
ASV2862	Erysipelotrichaceae	Bulleidia	NA
ASV2863	Lachnospiraceae	Catonella	NA
ASV2864	Muribaculaceae	NA	NA
ASV2865	Bacteroidaceae	Bacteroides	pyogenes
ASV2866	Caulobacteraceae	NA	NA
ASV2867	Lachnospiraceae	Oribacterium	NA
ASV2868	Burkholderiaceae	Curvibacter	NA
ASV2869	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2870	Prevotellaceae	Prevotella_2	NA

ASV2871	Mycoplasmataceae	Mycoplasma	NA
ASV2872	Flavobacteriaceae	Capnocytophaga	NA
ASV2873	Moraxellaceae	Acinetobacter	baumannii/calcoaceticus/nosocomialis/radioresistens
ASV2874	Lachnospiraceae	Butyrivibrio_2	NA
ASV2875	Spirochaetaceae	Treponema_2	NA
ASV2876	Acidobacteriaceae_(Subgroup_1)	Granulicella	NA
ASV2877	Spirochaetaceae	Treponema_2	NA
ASV2878	Lachnospiraceae	Johnsonella	NA
ASV2879	Leuconostocaceae	Leuconostoc	citreum/mesenteroides/pseudomesenteroides
ASV2880	Sphingomonadaceae	Novosphingobium	NA
ASV2881	Veillonellaceae	NA	NA
ASV2882	Neisseriaceae	NA	NA
ASV2883	NA	NA	NA
ASV2884	Lactobacillaceae	Lactobacillus	animalis/apodemi/crispatus/faecis/murinus
ASV2885	Nocardiaceae	Rhodococcus	baikonurensis/boritolerans/degradans/erythropolis/globerulus/hoagii/opacus/qingshengii/rhodochrous
ASV2886	Xanthomonadaceae	Stenotrophomonas	acidaminiphila/humi/maltophilia/pavanii
ASV2887	Lachnospiraceae	Catonella	NA
ASV2888	Family_XI	Anaerococcus	NA
ASV2889	Prevotellaceae	Prevotella	NA
ASV2890	Desulfomicrobiaceae	Desulfomicrobium	NA
ASV2891	Neisseriaceae	Alysiella	NA
ASV2892	Actinomycetaceae	Actinomyces	NA
ASV2893	Veillonellaceae	Selenomonas	NA
ASV2894	Lachnospiraceae	Catonella	NA
ASV2895	Campylobacteraceae	Campylobacter	NA

ASV2896	Atopobiaceae	Atopobium	NA
ASV2897	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia	NA
ASV2898	Anaerolineaceae	Flexilinea	NA
ASV2899	Corynebacteriaceae	Corynebacterium_1	accolens/fastidiosum/macginleyi/segmentosum
ASV2900	Rhodocyclaceae	Propionivibrio	NA
ASV2901	Lachnospiraceae	Catonella	NA
ASV2902	Gemmatimonadaceae	NA	NA
ASV2903	Geodermatophilaceae	Blastococcus	NA
ASV2904	Cardiobacteriaceae	NA	NA
ASV2905	NA	NA	NA
ASV2906	Mitochondria	NA	NA
ASV2907	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV2908	Streptococcaceae	Streptococcus	pneumoniae
ASV2909	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV2910	Sphingobacteriaceae	Sphingobacterium	cellulitidis/daejeonense/lactis/mizutaii
ASV2911	Lachnospiraceae	NA	NA
ASV2912	Caulobacteraceae	Brevundimonas	kwangchunensis
ASV2913	Family_XIII	NA	NA
ASV2914	Neisseriaceae	Eikenella	corrodens
ASV2915	Lachnospiraceae	Blautia	NA
ASV2916	Lachnospiraceae	Anaerostipes	hadrus
ASV2917	Burkholderiaceae	Caenimonas	NA
ASV2918	Erysipelotrichaceae	Catenibacterium	mitsuokai
ASV2919	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV2920	NA	NA	NA

ASV2921	Pseudomonadaceae	Pseudomonas	deceptionensis/endophytica/fluorescens/fragi/helleri/lundensis/psychrophila/putida/syringae/taiwanensis/weihenstephanensis
ASV2922	Family_XI	Peptoniphilus	massiliensis/urinimassiliensis
ASV2923	Prevotellaceae	Alloprevotella	NA
ASV2924	Micrococcaceae	Kocuria	arsenatis/rhizophila/salsicia
ASV2925	Acidaminococcaceae	Phascolarctobacterium	succinatutens
ASV2926	Lachnospiraceae	Oribacterium	NA
ASV2927	Flavobacteriaceae	Capnocytophaga	NA
ASV2928	Chthoniobacteraceae	Chthoniobacter	NA
ASV2929	Prevotellaceae	Prevotella	NA
ASV2930	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2931	Salinisphaeraceae	Salinisphaera	NA
ASV2932	Vibrionaceae	Vibrio	algivorus/casei/litoralis/penaeicida/rumoiensis
ASV2933	Prevotellaceae	Prevotella_7	NA
ASV2934	Lachnospiraceae	Stomatobaculum	NA
ASV2935	Muribaculaceae	NA	NA
ASV2936	Staphylococcaceae	Staphylococcus	arlettae/aureus/cohnii/gallinarum/haemolyticus/nepalensis/saccharolyticus/saprophyticus/succinus/xylosus
ASV2937	Family_XI	Gemella	NA
ASV2938	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2939	Ruminococcaceae	Faecalibacterium	NA
ASV2940	Mycoplasmataceae	Mycoplasma	NA
ASV2941	Flavobacteriaceae	Flavobacterium	aquidurense/collinsii/frigidimaris/hydatis/johnsoniae/saccharophilum
ASV2942	Burkholderiaceae	Acidovorax	defluvii/temperans
ASV2943	Weeksellaceae	Bergeyella	NA
ASV2944	NA	NA	NA
ASV2945	Micrococcaceae	Kocuria	assamensis/palustris
ASV2946	Mycoplasmataceae	Mycoplasma	NA

ASV2947	Microbacteriaceae	Galbitalea	NA
ASV2948	Flavobacteriaceae	Capnocytophaga	NA
ASV2949	Bacteroidaceae	Bacteroides	denticanum/pyogenes
ASV2950	Veillonellaceae	Veillonella	NA
ASV2951	Peptostreptococcaceae	NA	NA
ASV2952	Lachnospiraceae	Roseburia	NA
ASV2953	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV2954	Saccharimonadaceae	NA	NA
ASV2955	Prevotellaceae	Prevotella	ihumii
ASV2956	Flavobacteriaceae	Capnocytophaga	NA
ASV2957	Burkholderiaceae	Achromobacter	NA
ASV2958	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV2959	Streptococcaceae	Streptococcus	NA
ASV2960	ML635J-40_aquatic_group	NA	NA
ASV2961	Saccharimonadaceae	NA	NA
ASV2962	Xanthobacteraceae	Tardiphaga	NA
ASV2963	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2964	Prevotellaceae	Prevotella	NA
ASV2965	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV2966	Prevotellaceae	Prevotella_2	NA
ASV2967	Lachnospiraceae	Lachnoclostridium	NA
ASV2968	Lachnospiraceae	Johnsonella	NA
ASV2969	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2970	Xanthobacteraceae	Bradyrhizobium	NA
ASV2971	Lactobacillaceae	Lactobacillus	NA
ASV2972	Pasteurellaceae	Necropsobacter	NA

ASV2973	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV2974	Prevotellaceae	Alloprevotella	NA
ASV2975	Geodermatophilaceae	Blastococcus	aggregatus/massiliensis
ASV2976	Coriobacteriaceae	Collinsella	intestinalis/stercoris
ASV2977	Moraxellaceae	Acinetobacter	harbinensis/lwoffii
ASV2978	NA	NA	NA
ASV2979	Prevotellaceae	NA	NA
ASV2980	NA	NA	NA
ASV2981	NA	NA	NA
ASV2982	Campylobacteraceae	Campylobacter	NA
ASV2983	NA	NA	NA
ASV2984	Prevotellaceae	Prevotella	bivia/denticola
ASV2985	Nocardiaceae	Nocardia	abscessus/asiatica/asteroides/farcinica/higoensis/kroppenstedtii/otitidiscaviarum/shimofusensis
ASV2986	Veillonellaceae	Selenomonas_4	NA
ASV2987	NA	NA	NA
ASV2988	Lactobacillaceae	Lactobacillus	NA
ASV2989	Moraxellaceae	Psychrobacter	namhaensis
ASV2990	Sphingomonadaceae	Sphingomonas	NA
ASV2991	Gemmatimonadaceae	Gemmatimonas	NA
ASV2992	Lachnospiraceae	Johnsonella	NA
ASV2993	Prevotellaceae	Prevotella_7	NA
ASV2994	Sphingomonadaceae	Sphingomonas	NA
ASV2995	Ruminococcaceae	Candidatus_Soleaferrea	NA
ASV2996	Synergistaceae	Pyramidobacter	piscolens
ASV2997	Lachnospiraceae	Johnsonella	NA
ASV2998	Spirochaetaceae	Treponema_2	NA

ASV2999	Streptococcaceae	Streptococcus	NA
ASV3000	Bacillaceae	Anoxybacillus	amylolyticus/ayderensis/beppuensis/bogrovensis/flavithermus/gonensis/kamchatkensis/kaynarcensis/kestanbolensis/mongoliensis
ASV3001	Micrococcaceae	Nesterenkonia	lacusekhoensis
ASV3002	Pasteurellaceae	NA	NA
ASV3003	Lactobacillaceae	Lactobacillus	reuteri
ASV3004	Porphyromonadaceae	Porphyromonas	NA
ASV3005	Family_XI	W5053	NA
ASV3006	Micrococcaceae	Kocuria	atrinae/carniphila/gwangalliensis/marina
ASV3007	Leptotrichiaceae	Leptotrichia	NA
ASV3008	Moraxellaceae	Acinetobacter	albensis/haemolyticus
ASV3009	Corynebacteriaceae	Corynebacterium_1	NA
ASV3010	NA	NA	NA
ASV3011	Saccharimonadaceae	NA	NA
ASV3012	Prevotellaceae	NA	NA
ASV3013	Lachnospiraceae	Johnsonella	NA
ASV3014	Xanthobacteraceae	Afipia	NA
ASV3015	Spirochaetaceae	Treponema_2	NA
ASV3016	Veillonellaceae	Selenomonas	NA
ASV3017	Moraxellaceae	Acinetobacter	NA
ASV3018	Neisseriaceae	Eikenella	NA
ASV3019	Bacteroidaceae	Bacteroides	fragilis/ovatus
ASV3020	Micrococcaceae	Pseudarthrobacter	NA
ASV3021	NA	NA	NA
ASV3022	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3023	Veillonellaceae	Selenomonas_4	NA
ASV3024	Leptotrichiaceae	Leptotrichia	NA

ASV3025	Veillonellaceae	NA	NA
ASV3026	Lachnospiraceae	NA	NA
ASV3027	Bifidobacteriaceae	Scardovia	NA
ASV3028	NA	NA	NA
ASV3029	Xanthomonadaceae	Stenotrophomonas	NA
ASV3030	Moraxellaceae	Moraxella	bovis/bovoculi/caprae/equi/lacunata
ASV3031	Fusobacteriaceae	Fusobacterium	NA
ASV3032	Streptococcaceae	Streptococcus	canis/dysgalactiae/pyogenes
ASV3033	Veillonellaceae	Dialister	NA
ASV3034	NA	NA	NA
ASV3035	Prevotellaceae	NA	NA
ASV3036	Flavobacteriaceae	Capnocytophaga	NA
ASV3037	Lachnospiraceae	NA	NA
ASV3038	Prevotellaceae	Prevotella	NA
ASV3039	Mycoplasmataceae	Mycoplasma	NA
ASV3040	Micrococcaceae	Rothia	NA
ASV3041	Xanthomonadaceae	Stenotrophomonas	maltophilia/pavanii
ASV3042	Moraxellaceae	Alkanindiges	illinoisensis
ASV3043	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3044	Weeksellaceae	NA	NA
ASV3045	Rhizobiaceae	Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium	NA
ASV3046	Desulfovibrionaceae	Desulfovibrio	NA
ASV3047	Sphingomonadaceae	Blastomonas	natatoria/ursincola
ASV3048	Clostridiaceae_1	Clostridium_sensu_stricto_1	NA
ASV3049	Deinococcaceae	Deinococcus	actinosclerus/arenae/grandis/radiotolerans/soli/xianganensis

ASV3050	Micrococcaceae	Rothia	arfidiae/endophytica/marina/mucilaginosa/nasimurium
ASV3051	Burkholderiaceae	NA	NA
ASV3052	Flavobacteriaceae	Flavobacterium	NA
ASV3053	Family_XI	Peptoniphilus	gorbachii/harei/phoceensis/rhinitidis/senegalensis/tyrrelliae
ASV3054	Mycoplasmataceae	Mycoplasma	NA
ASV3055	Burkholderiaceae	Hydrogenophaga	atypica/bisanensis/defluvii/flava/palleronii/pseudoflava/taeniospiralis
ASV3056	Listeriaceae	Brochothrix	thermosphacta
ASV3057	Spirochaetaceae	Treponema_2	NA
ASV3058	Hymenobacteraceae	Hymenobacter	NA
ASV3059	Iamiaceae	Iamia	NA
ASV3060	Sphingomonadaceae	Sphingomonas	alpina/aquatica/aquatilis/insulae/kyungheensis/melonis/sanxanigenens/taxi
ASV3061	Beijerinckiaceae	Methylobacterium	bullatum/marchantiae
ASV3062	Streptococcaceae	Streptococcus	equi
ASV3063	Caulobacteraceae	Brevundimonas	diminuta/terrae
ASV3064	Mycobacteriaceae	Mycobacterium	chelonae/insubricum/llatzerense/mucogenicum/phlei/phocaicum/tuberculosis
ASV3065	Prevotellaceae	Prevotella	NA
ASV3066	Mycoplasmataceae	Mycoplasma	NA
ASV3067	Bifidobacteriaceae	Bifidobacterium	NA
ASV3068	Actinomycetaceae	Actinomyces	NA
ASV3069	Actinomycetaceae	Actinomyces	NA
ASV3070	Pasteurellaceae	NA	NA
ASV3071	Weeksellaceae	Chryseobacterium	anthropi/haifense
ASV3072	Prevotellaceae	Prevotella_7	NA
ASV3073	Family_XIII	NA	NA
ASV3074	Flavobacteriaceae	Flavobacterium	NA
ASV3075	Prevotellaceae	Prevotella_7	NA

ASV3076	Burkholderiaceae	Polaromonas	NA
ASV3077	Veillonellaceae	Selenomonas_4	NA
ASV3078	Lachnospiraceae	Stomatobaculum	NA
ASV3079	NA	NA	NA
ASV3080	NA	NA	NA
ASV3081	Bacillaceae	Geobacillus	kaustophilus/stearothermophilus
ASV3082	Mitochondria	NA	NA
ASV3083	Prevotellaceae	Prevotella_7	NA
ASV3084	NA	NA	NA
ASV3085	Prevotellaceae	Prevotella_7	NA
ASV3086	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3087	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV3088	Leptotrichiaceae	Leptotrichia	NA
ASV3089	Neisseriaceae	NA	NA
ASV3090	Defluviitaleaceae	Defluviitaleaceae_UCG-011	NA
ASV3091	Blastocatellaceae	NA	NA
ASV3092	Leptotrichiaceae	Leptotrichia	NA
ASV3093	Pedosphaeraceae	NA	NA
ASV3094	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3095	Spirochaetaceae	Treponema_2	NA
ASV3096	Saccharimonadaceae	NA	NA
ASV3097	Sphingobacteriaceae	Sphingobacterium	anhuiense/faecium/kitahiroshimense
ASV3098	Caulobacteraceae	Brevundimonas	staleyi/subvibrioides
ASV3099	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV3100	Veillonellaceae	NA	NA
ASV3101	Pasteurellaceae	NA	NA

ASV3102	Dermabacteraceae	Brachybacterium	alimentarium/conglomeratum/faecium/massiliense/paraconglomeratum/sacelli
ASV3103	NA	NA	NA
ASV3104	NA	NA	NA
ASV3105	NA	NA	NA
ASV3106	NA	NA	NA
ASV3107	NA	NA	NA
ASV3108	Prevotellaceae	Alloprevotella	NA
ASV3109	Veillonellaceae	Dialister	NA
ASV3110	Streptococcaceae	Streptococcus	NA
ASV3111	Streptococcaceae	Streptococcus	NA
ASV3112	Family_XIII	NA	NA
ASV3113	Prevotellaceae	Prevotella_7	NA
ASV3114	NA	NA	NA
ASV3115	Prevotellaceae	Alloprevotella	NA
ASV3116	Peptostreptococcaceae	Peptostreptococcus	NA
ASV3117	NA	NA	NA
ASV3118	Leptotrichiaceae	Leptotrichia	NA
ASV3119	Moraxellaceae	Alkanindiges	NA
ASV3120	Neisseriaceae	Bergeriella	NA
ASV3121	Family_XI	Gemella	NA
ASV3122	Corynebacteriaceae	Corynebacterium_1	urealyticum
ASV3123	Flavobacteriaceae	Flavobacterium	lindanitolerans
ASV3124	Alteromonadaceae	Rheinheimera	NA
ASV3125	Moraxellaceae	Acinetobacter	NA
ASV3126	Bifidobacteriaceae	Alloscardovia	NA
ASV3127	Spirochaetaceae	Treponema_2	NA

ASV3128	NA	NA	NA
ASV3129	Family_XIII	NA	NA
ASV3130	NA	NA	NA
ASV3131	Peptostreptococcaceae	NA	NA
ASV3132	Pasteurellaceae	Rodentibacter	NA
ASV3133	Flavobacteriaceae	Capnocytophaga	NA
ASV3134	Succinivibrionaceae	NA	NA
ASV3135	Veillonellaceae	NA	NA
ASV3136	Corynebacteriaceae	Corynebacterium_1	afermentans/coyleae/fournierii/H-EH3/ihumii/mucifaciens/pilbarense/ureicelerivorans
ASV3137	Saccharimonadaceae	NA	NA
ASV3138	Lachnospiraceae	Johnsonella	NA
ASV3139	Flavobacteriaceae	NA	NA
ASV3140	Dietziaceae	Dietzia	cercidiphylli/dagingensis/natronolimnaea/psychralcaliphila
ASV3141	Campylobacteraceae	Campylobacter	NA
ASV3142	Veillonellaceae	NA	NA
ASV3143	Solirubrobacteraceae	Solirubrobacter	NA
ASV3144	NA	NA	NA
ASV3145	Prevotellaceae	Prevotella	NA
ASV3146	Lachnospiraceae	Johnsonella	NA
ASV3147	NA	NA	NA
ASV3148	Lachnospiraceae	Moryella	NA
ASV3149	Synergistaceae	Fretibacterium	NA
ASV3150	Leptotrichiaceae	Sneathia	sanguinegens
ASV3151	Veillonellaceae	Negativicoccus	succinicivorans
ASV3152	Enterobacteriaceae	Hafnia-Obesumbacterium	NA
ASV3153	Weeksellaceae	Bergeyella	NA

ASV3154	NA	NA	NA
ASV3155	Veillonellaceae	Centipeda	NA
ASV3156	Family_XI	Parvimonas	NA
ASV3157	Burkholderiaceae	Comamonas	testosteroni/thiooxydans
ASV3158	Corynebacteriaceae	Corynebacterium	variabile
ASV3159	Family_XI	Finegoldia	magna
ASV3160	Burkholderiaceae	NA	NA
ASV3161	Spirochaetaceae	Treponema_2	NA
ASV3162	Peptostreptococcaceae	Peptostreptococcus	NA
ASV3163	Prevotellaceae	Prevotella_2	NA
ASV3164	Lachnospiraceae	NA	NA
ASV3165	Brevibacteriaceae	Brevibacterium	ammoniilyticum/aureum/casei/celere/epidermidis/iodinum/linens/permense/picturae/sandarakinum/sanguinis/sediminis/siliguriense
ASV3166	Spirosomaceae	Leadbetterella	NA
ASV3167	Weeksellaceae	Chishuiella	NA
ASV3168	Pseudomonadaceae	Pseudomonas	stutzeri/xanthomarina
ASV3169	Sphingomonadaceae	Rhizorhapis	NA
ASV3170	Rhodocyclaceae	Azonexus	fungiphilus
ASV3171	Lachnospiraceae	Agathobacter	NA
ASV3172	Family_XI	Ezakiella	massiliensis/peruensis
ASV3173	Hydrogenophilaceae	Hydrogenophilus	NA
ASV3174	Mycoplasmataceae	Mycoplasma	NA
ASV3175	Lachnospiraceae	Blautia	NA
ASV3176	Fusobacteriaceae	NA	NA
ASV3177	Burkholderiaceae	NA	NA
ASV3178	NA	NA	NA
ASV3179	Moraxellaceae	Acinetobacter	NA

ASV3180	NA	NA	NA
ASV3181	NA	NA	NA
ASV3182	Leptotrichiaceae	Leptotrichia	NA
ASV3183	Saccharimonadaceae	NA	NA
ASV3184	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3185	Erysipelotrichaceae	Erysipelatoclostridium	NA
ASV3186	NA	NA	NA
ASV3187	Rhizobiaceae	Ochrobactrum	intermedium/pituitosum/rhizosphaerae
ASV3188	Burkholderiaceae	Ralstonia	NA
ASV3189	Deinococcaceae	Deinococcus	radiopugnans/swuensis
ASV3190	Lachnospiraceae	NA	NA
ASV3191	Solirubrobacteraceae	Parviterribacter	NA
ASV3192	Weeksellaceae	Bergeyella	NA
ASV3193	Spirochaetaceae	Treponema_2	NA
ASV3194	Flavobacteriaceae	Capnocytophaga	NA
ASV3195	Mitochondria	NA	NA
ASV3196	Neisseriaceae	NA	NA
ASV3197	Pasteurellaceae	Cricetibacter	NA
ASV3198	Flavobacteriaceae	Capnocytophaga	NA
ASV3199	Lachnospiraceae	Johnsonella	NA
ASV3200	Prevotellaceae	Prevotella	NA
ASV3201	Porphyromonadaceae	Porphyromonas	NA
ASV3202	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3203	Enterobacteriaceae	NA	NA
ASV3204	Nocardioidaceae	Nocardioides	NA
ASV3205	Pasteurellaceae	NA	NA

ASV3206	Leptotrichiaceae	Oceanivirga	NA
ASV3207	Rhodobacteraceae	Haematobacter	massiliensis/missouriensis
ASV3208	Spirochaetaceae	Treponema_2	NA
ASV3209	Lactobacillaceae	Lactobacillus	NA
ASV3210	Lachnospiraceae	Stomatobaculum	NA
ASV3211	Leptotrichiaceae	Leptotrichia	NA
ASV3212	NA	NA	NA
ASV3213	Lachnospiraceae	Lachnoclostridium	NA
ASV3214	Lachnospiraceae	Butyrivibrio_2	NA
ASV3215	NA	NA	NA
ASV3216	Flavobacteriaceae	Capnocytophaga	NA
ASV3217	Pasteurellaceae	NA	NA
ASV3218	Brevibacteriaceae	Brevibacterium	ravenspurgense
ASV3219	Ruminococcaceae	Faecalibacterium	CM04-06
ASV3220	Leuconostocaceae	Leuconostoc	carnosum/citreum/garlicum/gelidum/holzapfelii/lactis/mesenteroides/palmae/pseudomesenteroides
ASV3221	Neisseriaceae	Bergeriella	NA
ASV3222	Prevotellaceae	Prevotella_7	NA
ASV3223	Lachnospiraceae	Blautia	NA
ASV3224	Lachnospiraceae	NA	NA
ASV3225	Bifidobacteriaceae	Bifidobacterium	crudilactis/psychraerophilum
ASV3226	Porphyromonadaceae	Porphyromonas	somerae
ASV3227	Prevotellaceae	Prevotella_7	NA
ASV3228	Caulobacteraceae	Brevundimonas	NA
ASV3229	Flavobacteriaceae	Capnocytophaga	NA
ASV3230	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV3231	Desulfobulbaceae	Desulfobulbus	NA

ASV3232	NA	NA	NA
ASV3233	Ruminococcaceae	UBA1819	NA
ASV3234	Fimbriimonadaceae	NA	NA
ASV3235	Bifidobacteriaceae	Scardovia	NA
ASV3236	NA	NA	NA
ASV3237	NA	NA	NA
ASV3238	Streptococcaceae	Lactococcus	piscium/plantarum
ASV3239	Pseudomonadaceae	Pseudomonas	japonica/putida
ASV3240	NA	NA	NA
ASV3241	Prevotellaceae	Alloprevotella	NA
ASV3242	Aerococcaceae	Aerococcus	urinaeequi/viridans
ASV3243	Campylobacteraceae	Campylobacter	NA
ASV3244	Lachnospiraceae	Moryella	NA
ASV3245	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia	NA
ASV3246	Peptostreptococcaceae	NA	NA
ASV3247	Family_XI	Anaerococcus	vaginalis
ASV3248	Lachnospiraceae	Blautia	NA
ASV3249	Microbacteriaceae	Frigoribacterium	faeni
ASV3250	Rikenellaceae	Blvii28_wastewater-sludge_group	NA
ASV3251	NA	NA	NA
ASV3252	Pasteurellaceae	NA	NA
ASV3253	Rhodocyclaceae	Methyloversatilis	discipulorum/universalis
ASV3254	NA	NA	NA
ASV3255	Burkholderiaceae	NA	NA
ASV3256	Campylobacteraceae	Campylobacter	NA

ASV3257	NA	NA	NA
ASV3258	NA	NA	NA
ASV3259	Atopobiaceae	Atopobium	NA
ASV3260	Family_XI	Gemella	NA
ASV3261	Tannerellaceae	NA	NA
ASV3262	Lachnospiraceae	Johnsonella	NA
ASV3263	Campylobacteraceae	Campylobacter	NA
ASV3264	Moraxellaceae	Acinetobacter	baumannii/junii/seohaensis/soli/towneri/venetianus
ASV3265	Fusobacteriaceae	NA	NA
ASV3266	Streptococcaceae	Streptococcus	NA
ASV3267	Saccharimonadaceae	NA	NA
ASV3268	Peptostreptococcaceae	NA	NA
ASV3269	Veillonellaceae	Centipeda	NA
ASV3270	NA	NA	NA
ASV3271	Xanthomonadaceae	Luteimonas	NA
ASV3272	Micrococcaceae	Kocuria	kristinae/rosea
ASV3273	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3274	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3275	Leptotrichiaceae	Leptotrichia	NA
ASV3276	Cellulomonadaceae	Tropheryma	whipplei
ASV3277	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV3278	Bacillaceae	NA	NA
ASV3279	NA	NA	NA
ASV3280	Geodermatophilaceae	Modestobacter	marinus/muralis/versicolor
ASV3281	Lachnospiraceae	NA	NA
ASV3282	Spirochaetaceae	Treponema_2	NA

ASV3283	Prevotellaceae	NA	NA
ASV3284	Spirochaetaceae	Treponema_2	NA
ASV3285	Prevotellaceae	Prevotella_6	bergensis
ASV3286	Family_XI	Gallicola	NA
ASV3287	Prevotellaceae	Prevotella	NA
ASV3288	Bacteroidaceae	Bacteroides	faecichinchillae/faecis/finegoldii/thetaiotaomicron
ASV3289	Lactobacillaceae	Lactobacillus	NA
ASV3290	Family_XIII	Mogibacterium	NA
ASV3291	Erysipelotrichaceae	Eggerthia	NA
ASV3292	Crocinitomicaceae	Fluviicola	NA
ASV3293	Pseudomonadaceae	Pseudomonas	argentinensis/arsenicoxydans/baetica/borealis/brassicacearum/brenneri/chlororaphis/corrugata/denitrificans/ficuserectae/fluorescens
ASV3294	Sphingomonadaceae	Sphingomonas	NA
ASV3295	NA	NA	NA
ASV3296	Paludibacteraceae	F0058	NA
ASV3297	Flavobacteriaceae	Capnocytophaga	NA
ASV3298	Saccharimonadaceae	NA	NA
ASV3299	Bacteroidales_Incertae_Sedis	Phocaeicola	NA
ASV3300	Enterobacteriaceae	Buttiauxella	brennerae/gaviniae
ASV3301	Sphingomonadaceae	Sphingobium	NA
ASV3302	Family_XII	Exiguobacterium	acetylicum/aestuarii/alkaliphilum/aquaticum/arabatum/aurantiacum/chiriqhucha/himgiriensis/homiense/marinum/mexicanum
ASV3303	Nocardioidaceae	Nocardioides	rubriscoriae
ASV3304	NA	NA	NA
ASV3305	Family_XI	Parvimonas	NA
ASV3306	Prevotellaceae	Prevotella_7	NA
ASV3307	Sphingomonadaceae	Sphingomonas	NA
ASV3308	Weeksellaceae	Chryseobacterium	anthropi

ASV3309	Family_XI	Parvimonas	NA
ASV3310	Erysipelotrichaceae	Eggerthia	NA
ASV3311	Rhodothermaceae	NA	NA
ASV3312	NA	NA	NA
ASV3313	Family_XIII	NA	NA
ASV3314	Prevotellaceae	Prevotella	NA
ASV3315	p-2534-18B5_gut_group	NA	NA
ASV3316	Prevotellaceae	Alloprevotella	NA
ASV3317	Lachnospiraceae	NA	NA
ASV3318	Prevotellaceae	Prevotella_2	NA
ASV3319	Caulobacteraceae	Phenylobacterium	muchangponense
ASV3320	Burkholderiaceae	Duganella	phyllosphaerae/radicis/sacchari/zoogloeoides
ASV3321	NA	NA	NA
ASV3322	Burkholderiaceae	Noviherbaspirillum	NA
ASV3323	Corynebacteriaceae	NA	NA
ASV3324	Ruminococcaceae	Ruminococcus_2	NA
ASV3325	NA	NA	NA
ASV3326	Prevotellaceae	Alloprevotella	NA
ASV3327	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3328	Weeksellaceae	NA	NA
ASV3329	Family_XI	Anaerococcus	nagyae
ASV3330	Corynebacteriaceae	Corynebacterium_1	jeddahense
ASV3331	Lachnospiraceae	Shuttleworthia	NA
ASV3332	Lachnospiraceae	Johnsonella	NA
ASV3333	Weeksellaceae	Bergeyella	NA
ASV3334	Eggerthellaceae	Cryptobacterium	NA

ASV3335	Sphingomonadaceae	Sphingomonas	NA
ASV3336	Flavobacteriaceae	Capnocytophaga	NA
ASV3337	Chitinophagaceae	Sediminibacterium	salmoneum
ASV3338	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV3339	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV3340	NA	NA	NA
ASV3341	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3342	Sphingobacteriaceae	Sphingobacterium	NA
ASV3343	Desulfobulbaceae	Desulfobulbus	NA
ASV3344	NA	NA	NA
ASV3345	Cellvibrionaceae	NA	NA
ASV3346	Prevotellaceae	Alloprevotella	NA
ASV3347	Family_XI	Ezakiella	NA
ASV3348	Veillonellaceae	NA	NA
ASV3349	NA	NA	NA
ASV3350	Prevotellaceae	Prevotella	NA
ASV3351	NA	NA	NA
ASV3352	Ruminococcaceae	Fastidiosipila	NA
ASV3353	Fusobacteriaceae	Fusobacterium	NA
ASV3354	NA	NA	NA
ASV3355	Flavobacteriaceae	Capnocytophaga	NA
ASV3356	Sphingomonadaceae	Sphingobium	yanoikuyae
ASV3357	Saccharimonadaceae	NA	NA
ASV3358	Acidaminococcaceae	Acidaminococcus	intestini
ASV3359	Cardiobacteriaceae	Cardiobacterium	NA
ASV3360	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA

ASV3361	Nocardiaceae	Williamsia	faeni/limnetica/muralis
ASV3362	Sphingomonadaceae	Sphingobium	japonicum/lactosutens/xenophagum
ASV3363	Deinococcaceae	Deinococcus	radiodurans
ASV3364	Rhizobiaceae	Neorhizobium	NA
ASV3365	Endomicrobiaceae	Candidatus_Endomicrobium	NA
ASV3366	Neisseriaceae	NA	NA
ASV3367	Campylobacteraceae	Campylobacter	NA
ASV3368	NA	NA	NA
ASV3369	NA	NA	NA
ASV3370	NA	NA	NA
ASV3371	Porphyromonadaceae	Porphyromonas	NA
ASV3372	Lactobacillaceae	Lactobacillus	fermentum/mucosae
ASV3373	Pasteurellaceae	Cricetibacter	NA
ASV3374	Prevotellaceae	NA	NA
ASV3375	NA	NA	NA
ASV3376	Family_XIII	NA	NA
ASV3377	Solirubrobacteraceae	Solirubrobacter	NA
ASV3378	Xanthomonadaceae	Arenimonas	daechungensis
ASV3379	Prevotellaceae	Prevotella	NA
ASV3380	Lachnospiraceae	NA	NA
ASV3381	NA	NA	NA
ASV3382	Family_XIII	NA	NA
ASV3383	Corynebacteriaceae	Corynebacterium_1	NA
ASV3384	Veillonellaceae	Selenomonas_3	NA
ASV3385	NA	NA	NA
ASV3386	Rhodanobacteraceae	Tahibacter	aquaticus

ASV3387	Sphingobacteriaceae	Pedobacter	insulae
ASV3388	Xanthomonadaceae	Pseudoxanthomonas	japonensis/mexicana
ASV3389	Prevotellaceae	Prevotella	NA
ASV3390	Spirochaetaceae	Treponema_2	NA
ASV3391	NA	NA	NA
ASV3392	NA	NA	NA
ASV3393	Ruminococcaceae	NA	NA
ASV3394	Porphyromonadaceae	Porphyromonas	NA
ASV3395	Ruminococcaceae	Fastidiosipila	NA
ASV3396	Family_XI	Parvimonas	NA
ASV3397	Family_XIII	Anaerovorax	NA
ASV3398	Veillonellaceae	NA	NA
ASV3399	p-2534-18B5_gut_group	NA	NA
ASV3400	Prevotellaceae	Prevotella	NA
ASV3401	Lachnospiraceae	Howardella	NA
ASV3402	Family_XI	Gemella	NA
ASV3403	Burkholderiaceae	Sutterella	NA
ASV3404	Fusobacteriaceae	Fusobacterium	NA
ASV3405	Lachnospiraceae	Stomatobaculum	NA
ASV3406	Rhodobacteraceae	Paracoccus	laeviglucosivorans/yeei
ASV3407	NA	NA	NA
ASV3408	Pasteurellaceae	NA	NA
ASV3409	Lactobacillaceae	Lactobacillus	curvatus/graminis/plantarum/sakei
ASV3410	Neisseriaceae	Bergeriella	NA
ASV3411	Leptotrichiaceae	Leptotrichia	NA
ASV3412	NA	NA	NA

ASV3413	Bifidobacteriaceae	Bifidobacterium	NA
ASV3414	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV3415	Porphyromonadaceae	Porphyromonas	NA
ASV3416	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV3417	Peptostreptococcaceae	Peptostreptococcus	NA
ASV3418	Tannerellaceae	Tannerella	NA
ASV3419	Family_XIII	Family_XIII_UCG-001	NA
ASV3420	Fusobacteriaceae	Fusobacterium	NA
ASV3421	Prevotellaceae	NA	NA
ASV3422	Saccharimonadaceae	NA	NA
ASV3423	NA	NA	NA
ASV3424	Prevotellaceae	Alloprevotella	NA
ASV3425	Pseudomonadaceae	Pseudomonas	NA
ASV3426	Rubrobacteriaceae	Rubrobacter	NA
ASV3427	Prevotellaceae	NA	NA
ASV3428	NA	NA	NA
ASV3429	Porphyromonadaceae	Porphyromonas	NA
ASV3430	Erysipelotrichaceae	NA	NA
ASV3431	Aerococcaceae	Aerococcus	NA
ASV3432	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3433	Family_XI	Ezakiella	NA
ASV3434	Family_XIII	S5-A14a	NA
ASV3435	Veillonellaceae	Centipeda	NA
ASV3436	Pasteurellaceae	NA	NA
ASV3437	NA	NA	NA
ASV3438	Fusobacteriaceae	NA	NA

ASV3439	Prevotellaceae	Prevotella_7	NA
ASV3440	Ruminococcaceae	Ruminococcaceae_UCG-014	NA
ASV3441	Prevotellaceae	NA	NA
ASV3442	Bacteriovoracaceae	Bacteriovorax	NA
ASV3443	Sphingobacteriaceae	Pedobacter	NA
ASV3444	Beijerinckiaceae	Bosea	eneae/lathyri/massiliensis/thiooxidans/vaviloviae/vestrisii
ASV3445	Lachnospiraceae	Stomatobaculum	NA
ASV3446	Porphyromonadaceae	Porphyromonas	NA
ASV3447	Campylobacteraceae	Campylobacter	NA
ASV3448	Porphyromonadaceae	Porphyromonas	asaccharolytica
ASV3449	Paludibacteraceae	F0058	NA
ASV3450	Methylophilaceae	Methylotenera	NA
ASV3451	Prevotellaceae	Alloprevotella	NA
ASV3452	Burkholderiaceae	Massilia	arvi/pinisoli/putida
ASV3453	Peptococcaceae	Peptococcus	NA
ASV3454	NA	NA	NA
ASV3455	Desulfovibrionaceae	Desulfovibrio	desulfuricans/multispirans
ASV3456	Lactobacillaceae	Lactobacillus	mali/nagelii/satsumensis/sicerae
ASV3457	Erysipelotrichaceae	Eggerthia	NA
ASV3458	Lachnospiraceae	Oribacterium	NA
ASV3459	Propionibacteriaceae	Pseudopropionibacterium	NA
ASV3460	Lactobacillaceae	Lactobacillus	NA
ASV3461	Staphylococcaceae	Staphylococcus	auricularis/epidermidis/warneri
ASV3462	NA	NA	NA
ASV3463	Lactobacillaceae	Lactobacillus	NA

ASV3464	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia	NA
ASV3465	Mitochondria	NA	NA
ASV3466	Prevotellaceae	Prevotella_2	NA
ASV3467	Sphingomonadaceae	Blastomonas	aquatica
ASV3468	Pseudomonadaceae	Pseudomonas	alcaligenes/alkylphenolica/cichorii/donghuensis/filiscindens/japonica/lundensis/multiaromavorans/plecoglossicida/putida/vranovensis
ASV3469	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3470	Spirosomaceae	Rudanella	NA
ASV3471	NA	NA	NA
ASV3472	Family_XI	Peptoniphilus	NA
ASV3473	Erysipelotrichaceae	Erysipelotrichaceae_UCG-006	NA
ASV3474	Eggerthellaceae	Cryptobacterium	NA
ASV3475	Family_XI	Peptoniphilus	NA
ASV3476	Acidaminococcaceae	Phascolarctobacterium	NA
ASV3477	Mycobacteriaceae	Mycobacterium	NA
ASV3478	Azospirillaceae	Skermanella	aerolata
ASV3479	Flavobacteriaceae	Capnocytophaga	NA
ASV3480	Eggerthellaceae	Slackia	NA
ASV3481	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3482	Lachnospiraceae	Howardella	NA
ASV3483	NA	NA	NA
ASV3484	Xanthomonadaceae	Stenotrophomonas	NA
ASV3485	NA	NA	NA
ASV3486	TRA3-20	NA	NA
ASV3487	Chthoniobacteraceae	Candidatus_Udaeobacter	NA
ASV3488	Leuconostocaceae	Leuconostoc	gelidum/inhae

ASV3489	Erysipelotrichaceae	Erysipelatoclostridium	ramosum
ASV3490	Prevotellaceae	NA	NA
ASV3491	Veillonellaceae	Selenomonas_3	NA
ASV3492	Spirochaetaceae	NA	NA
ASV3493	NA	NA	NA
ASV3494	Veillonellaceae	Dialister	NA
ASV3495	Burkholderiaceae	NA	NA
ASV3496	Lachnospiraceae	Stomatobaculum	NA
ASV3497	Lachnospiraceae	Stomatobaculum	NA
ASV3498	Carnobacteriaceae	NA	NA
ASV3499	Bacteroidales_Incertae_Sedis	Phocaeicola	NA
ASV3500	Rhodobacteraceae	Paracoccus	NA
ASV3501	NA	NA	NA
ASV3502	Family_XI	Gemella	NA
ASV3503	Erysipelotrichaceae	Bulleidia	NA
ASV3504	NA	NA	NA
ASV3505	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV3506	Beijerinckiaceae	Methylobacterium	adhaesivum
ASV3507	Neisseriaceae	Bergeriella	NA
ASV3508	Neisseriaceae	NA	NA
ASV3509	Lachnospiraceae	Lachnoanaerobaculum	NA
ASV3510	Family_XI	Parvimonas	NA
ASV3511	Paenibacillaceae	Paenibacillus	NA
ASV3512	NA	NA	NA
ASV3513	Ruminococcaceae	Faecalibacterium	prausnitzii
ASV3514	Leptotrichiaceae	Leptotrichia	NA

ASV3515	NA	NA	NA
ASV3516	Spirosomaceae	Dyadobacter	NA
ASV3517	Sphingomonadaceae	Sphingopyxis	NA
ASV3518	Streptococcaceae	Streptococcus	NA
ASV3519	NA	NA	NA
ASV3520	Prevotellaceae	NA	NA
ASV3521	Prevotellaceae	Prevotella_6	corporis
ASV3522	Lachnospiraceae	Moryella	NA
ASV3523	Porphyromonadaceae	Porphyromonas	NA
ASV3524	Family_XI	Anaerococcus	NA
ASV3525	Prevotellaceae	Prevotella	NA
ASV3526	NA	NA	NA
ASV3527	Pseudomonadaceae	Pseudomonas	aeruginosa/citronellolis/delhiensis/denitrificans/knackmussii/nitritireducens/nitroreducens/panipatensis/stutzeri
ASV3528	Flavobacteriaceae	Flavobacterium	ummariense
ASV3529	Fusobacteriaceae	NA	NA
ASV3530	Xanthomonadaceae	Stenotrophomonas	acidaminiphila/maltophilia/nitritireducens
ASV3531	NA	NA	NA
ASV3532	Mitochondria	NA	NA
ASV3533	Neisseriaceae	NA	NA
ASV3534	Mycoplasmataceae	Mycoplasma	NA
ASV3535	NA	NA	NA
ASV3536	Atopobiaceae	Atopobium	NA
ASV3537	Ruminococcaceae	NA	NA
ASV3538	Mitochondria	NA	NA
ASV3539	Family_XI	Anaerococcus	mediterraneensis/murdochii
ASV3540	Prevotellaceae	Prevotella	buccalis

ASV3541	NA	NA	NA
ASV3542	Reyranellaceae	Reyranella	soli
ASV3543	Mycoplasmataceae	Mycoplasma	NA
ASV3544	NA	NA	NA
ASV3545	NA	NA	NA
ASV3546	Xanthobacteraceae	Afipia	clevelandensis
ASV3547	Prevotellaceae	Alloprevotella	NA
ASV3548	Actinomycetaceae	F0332	NA
ASV3549	NA	NA	NA
ASV3550	Lachnospiraceae	Stomatobaculum	NA
ASV3551	Pasteurellaceae	NA	NA
ASV3552	Erysipelotrichaceae	NA	NA
ASV3553	Campylobacteraceae	Campylobacter	NA
ASV3554	NA	NA	NA
ASV3555	Burkholderiaceae	Massilia	aurea/brevitalea/cf./namucuonensis/timonae
ASV3556	Hymenobacteraceae	Hymenobacter	NA
ASV3557	Lachnospiraceae	Lachnospiraceae_NK3A20_group	NA
ASV3558	Veillonellaceae	Selenomonas_4	NA
ASV3559	Lachnospiraceae	Coprococcus_3	comes
ASV3560	Pasteurellaceae	NA	NA
ASV3561	NA	NA	NA
ASV3562	Methanobacteriaceae	Methanobrevibacter	NA
ASV3563	NA	NA	NA
ASV3564	Pseudoalteromonadaceae	Pseudoalteromonas	NA
ASV3565	Prevotellaceae	NA	NA
ASV3566	NA	NA	NA

ASV3567	Leptotrichiaceae	Leptotrichia	NA
ASV3568	Leptotrichiaceae	Leptotrichia	NA
ASV3569	NA	NA	NA
ASV3570	Weeksellaceae	Chryseobacterium	aahli/defluvii/limigenitum/piperi/soldanellicola/soli/yeoncheonense
ASV3571	env.OPS_17	NA	NA
ASV3572	NA	NA	NA
ASV3573	NA	NA	NA
ASV3574	Chitinophagaceae	NA	NA
ASV3575	Rikenellaceae	Rikenellaceae_RC9_gut_group	NA
ASV3576	Saccharimonadaceae	Candidatus_Saccharimonas	NA
ASV3577	Flavobacteriaceae	Capnocytophaga	NA
ASV3578	Prevotellaceae	NA	NA
ASV3579	NA	NA	NA
ASV3580	Family_XI	Peptoniphilus	NA
ASV3581	NA	NA	NA
ASV3582	Saccharimonadaceae	NA	NA
ASV3583	Lachnospiraceae	Lachnospiraceae_NK3A20_group	NA
ASV3584	Burkholderiaceae	Verticia	NA
ASV3585	NA	NA	NA
ASV3586	Prevotellaceae	Prevotellaceae_YAB2003_group	NA
ASV3587			

Appendix E: Chapter 7

This appendix contains additional relevant work completed after the study presented in Chapter 7. In Chapter 7 I have undertaken characterisation of the saliva microbiota of the polygenic risk score cohort investigated in the same chapter; here I have applied similar methods to the wider TwinsUK saliva microbiota sample to increase the sample size and thus power. Some of this work in relation to alpha diversity will not be submitted for publication as no association remained after adjustment for multiple testing. However, for brevity I have included these analyses in the thesis.

Alpha Diversity

	Shannon			Simpson			Observed		
	Estimate	**P**	**Q**	**Estimate**	**P**	**Q**	**Estimate**	**P**	**Q**
Age	0.00807	0.055	0.442	0.0093	0.025	0.198	-0.00676	0.097	0.195
Sex M	-0.23255	0.189	0.652	-0.2138	0.220	0.439	0.00310	0.986	0.997
BMI	0.00806	0.326	0.652	0.0136	0.102	0.375	0.00003	0.997	0.997
Diet	-0.00121	0.785	0.785	0.0018	0.684	0.684	-0.00387	0.363	0.580
Seq. Depth	0.00000	0.589	0.673	0.0000	0.568	0.649	0.000004	0.04	0.160
Not Fasted	-0.17434	0.315	0.652	-0.1686	0.335	0.533	-0.29523	0.076	0.196
Storage Time	0.00065	0.490	0.654	0.0008	0.400	0.533	0.001914	0.035	0.161

Supplementary table E 7.1 Alpha diversity of the saliva microbiota association of factors within the wider TwinsUK sample N= 529.

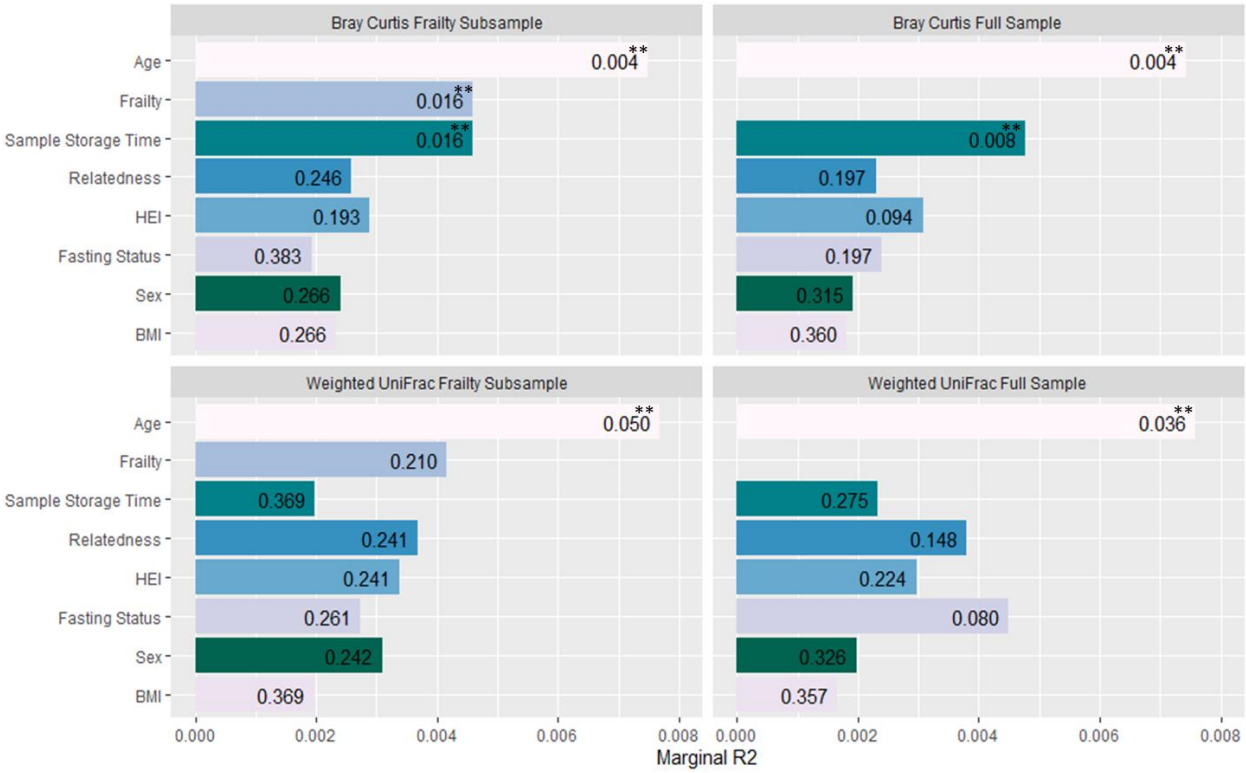
	Shannon			Simpson			Observed		
	Estimate	P	Q	Estimate	P	Q	Estimate	P	Q
Age	0.011525	0.012	0.107	0.01176	0.010	0.089	-0.0017	0.693	0.813
Sex M	-0.205960	0.275	0.63	-0.15826	0.399	0.695	-0.0538	0.766	0.813
BMI	0.009602	0.280	0.63	0.01639	0.073	0.218	0.0020	0.813	0.813
Diet	-0.002269	0.643	0.739	0.00096	0.848	0.848	-0.0034	0.456	0.684
Frailty Index	-0.676624	0.105	0.472	-0.76796	0.073	0.218	-0.6745	0.084	0.251
Seq. Depth	0.000000	0.840	0.84	0.00000	0.617	0.695	0.0000	0.019	0.086
Storage Time	0.000658	0.523	0.739	0.00063	0.541	0.695	0.0023	0.019	0.086
Not Fasted	-0.090284	0.657	0.739	-0.13859	0.505	0.695	-0.2367	0.214	0.41

Supplementary Table E 7.2. Alpha diversity of the saliva microbiota association of factors within the wider TwinsUK sample. Subsample analysis of participants for whom frailty were available N=511.

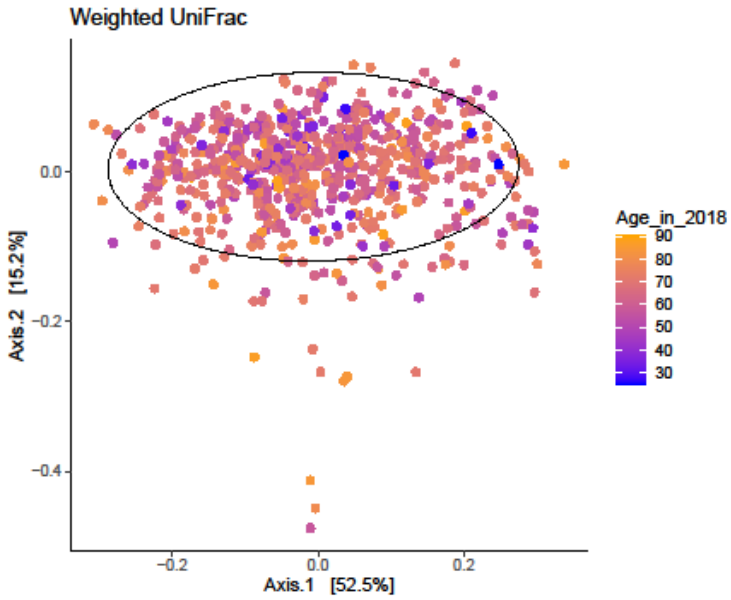
There were demographic differences between those with and without frailty data. Those missing frailty data had a median age of 51 (IQR 25 – 57), whilst those with frailty data had a median age of 67 (IQR 60 – 66). Participants who were missing frailty data were significantly younger (p = 1.08e-12).

There may be differences in the these younger participants which drive the lower alpha diversity association demonstrated in the full sample (Supplementary Table E 7.1) compared to the frailty subsample analysis (Supplementary Table E 7.2). To this end, differences in oral hygiene habits were investigated. However, only 180 participants of 689 had completed the oral hygiene questionnaire. Within this subsample of 180, there was no difference in tooth brushing habits between those with and without frailty data. Within both groups, 81% of participants brushed their teeth once a day, whilst the remainder brushed once a day.

Beta Diversity

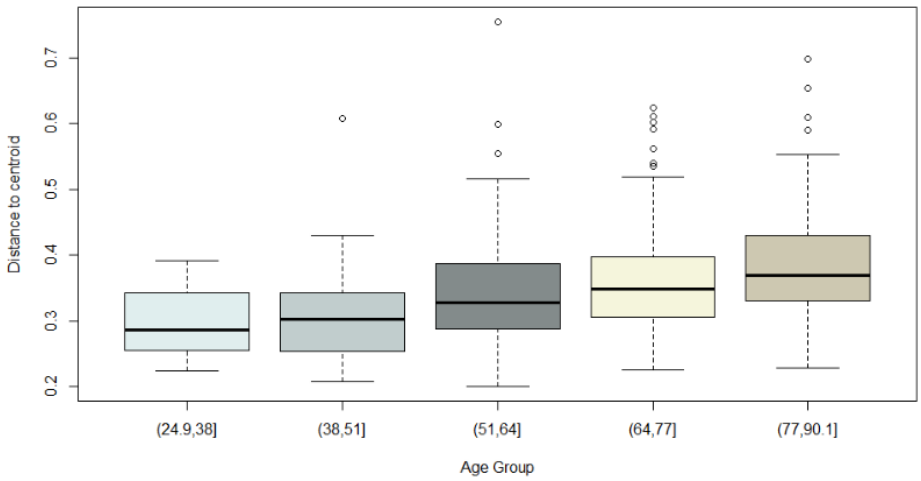


Supplementary Figure E 7.1 Association of factors with beta diversity of the saliva microbiota
Age was significantly associated with beta diversity after adjusting for all other factors listed plus sequencing depth (sample n = 529). The frailty subsample includes only participants for whom frailty data were available (n = 511). Bars are annotated with FDR adjusted p values (Q values).



Supplementary Figure E 7.2. PCoA ordination of Weighted UniFrac of the saliva microbiota in relation to age (n = 529). On ordination there was no clear clustering of age, however there was an indication that older participants tended to have lower beta diversity dissimilarity of the saliva microbiota than younger participants.

Across 5 age groups, variation in distribution of beta diversity increased sequentially with age, demonstrating that younger participants were more similar in their saliva microbiota diversity than older participants (Figure 3).



Supplementary Figure E 7.3 Bray Curtis beta dispersion of age within full cohort sample.
A difference in the variance in across age groups was demonstrated (5.125e-06, ANOVA).

A subsample analysis of participants for whom periodontal disease data were available (n=124) is presented in Supplementary Table E 7.3 . Periodontal disease was defined as presence of either bleeding gums or loose teeth, or having been told by a dentist that they have gum disease. These data were collected using self-report questionnaires. In the sub-sample analysis of participants with periodontal disease data, implementing a multivariate model as prior with the addition of periodontal disease, age was associated with Bray Curtis dissimilarity (Q = 0.04), but not Weighted Uni Frac (Q = 0.278). The loss of association of age with Weighted Unifrac in the periodontal disease subsample analysis may potentially be a power issue; the periodontal disease sub-sample was substantially smaller (n=124). Periodontal disease was associated with both Bray Curtis dissimilarity (Q = 0.01) and Weighted Uni Frac distance (Q = 0.03).

	Bray Curtis		Weighted UniFrac	
	R2	Q	R2	Q
Relatedness	0.013	0.364	0.005	0.627
Sample Storage Time	0.008	0.363	0.006	0.573
Fasting Status	0.009	0.281	0.011	0.295
BMI	0.010	0.208	0.008	0.497
HEI	0.011	0.140	0.011	0.295
Frailty	0.012	0.547	0.025	0.060
Sex	0.006	0.107	0.031	0.060
Periodontal Disease	0.016	0.01**	0.045	0.03**
Age	0.038	0.04**	0.016	0.278

Supplementary Table E 7.3 **Oral health subsample PERMANOVA including participants with data for periodontal disease (n=124).**